# ANALYSIS OF THE ROLES OF THE P10 DOMAIN OF GAG AND OF DIMERIZATION IN ROUS SARCOMA VIRUS ASSEMBLY

**A Dissertation**

**Presented to the Faculty of the Graduate School**

**of Cornell University**

**in Partial Fulfillment of the Requirements for the Degree of**

**Doctor of Philosophy**

**by**

**Judith Marie Phillips**

**January 2009**

# ANALYSIS OF THE ROLES OF THE P10 DOMAIN OF GAG AND OF DIMERIZATION IN ROUS SARCOMA VIRUS ASSEMBLY

**Judith Marie Phillips, D.V.M., Ph.D.**

**Cornell University 2009**

In the Rous sarcoma virus (RSV) Gag protein, the 25 amino acid residues of the p10 domain immediately upstream of the CA domain are essential for spherical immature particle formation. The significance of this finding was not known at the inception of this work. I performed a systematic mutagenesis on this region and found excellent correlation between the amino acid side chains required for *in vitro* assembly and those that participate in the p10-CA dimer interface in a previously described crystal structure. I then introduced exogenous cysteine residues that were predicted to form disulfide bonds across the dimer interface. Upon oxidation of immature particles a disulfide-linked Gag hexamer was formed, implying that p10 participates in and stabilizes the immature Gag hexamer. This is the first example of a critical interaction between two different Gag domains. Molecular modeling of the RSV immature hexamer performed by a collaborator, Paul Murray, indicates that the N-terminal domains of CA must expand relative to the modeled mature hexamer to accommodate the p10 contact.

Replacement of the NC domain of HIV-1 Gag with an artificial dimerization domain has been shown to support assembly of virus-like particles *in vitro* and in cells. Similar systems for RSV Gag have yielded inconsistent results, suggesting that the requirements for RSV assembly may

differ from those for HIV-1 assembly. I attempted to clarify these requirements in native avian cells and *in vitro* using chimeric Gag proteins in which NC was replaced by a leucine zipper domain or by an inducible dimerization domain based on the FK506-binding protein (FKBP). Budding of the Gag-leucine zipper fusion protein from avian cells was observed but was not entirely consistent, and similar fusion proteins failed to assemble *in vitro*. The Gag-FKBP fusion proteins similarly failed to respond to the induction agent. I concluded that further characterization of Gag dimerization and of the sequence upstream of NC will be required before an assembly system based on artificial dimerization can be developed for RSV Gag.

## BIOGRAPHICAL SKETCH

Judith Marie Phillips was born on May 26, 1979 in Ames, Iowa, an event that may have hastened the completion of her father's Ph.D. studies. She spent her childhood in Chester County, Pennsylvania, where she began her biology training by turning over rocks and identifying the creatures underneath. After graduating from Sanford School in Hockessin, Delaware, she attended Kenyon College, where she worked with Professor Joan L. Slonczewski on pH-inducible genes in *Escherichia coli* and *Helicobacter pylori* before embarking on a senior thesis project with Professor Elizabeth A. Ottinger on the binding of Src SH2 and SH3 domains to peptide analogs of the p130 Cas protein. Following her graduation from Kenyon in 2000, she thought that she had finished with research and entered the Cornell University College of Veterinary Medicine with the intention of going into private practice. The allure of the bench proved stronger than expected, and following a summer working with Dr. Kenneth Simpson on *Helicobacter* infection in cats as part of Cornell's Leadership Program for Veterinary Students, she enrolled in the newly-founded Dual D.V.M./Ph.D. program and joined the laboratory of Professor Volker M. Vogt to study retrovirus assembly. She completed the D.V.M. program in 2006. Later that summer, she married Karl F. Newman of Cortland, New York, and the completion of her Ph.D. studies is currently being hastened by the impending arrival of their first child.

## ACKNOWLEDGMENTS

I would like to thank:

My advisor, Volker Vogt, for taking a chance on (another) unorthodox student, for keeping the faith in my research when I'd lost hope, and for many interesting conversations

My labmates, for ideas, encouragement, and much needed constructive criticism, as well as for keeping life interesting

The Dual Degree Program, for an enriching academic experience as well as administrative and financial support

And my husband, family, and friends, for reminding me that there is life outside the lab

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

Viruses

| | |
|---|---|
| CMV | cytomegalovirus |
| EIAV | equine infectious anemia virus |
| HFV | human foamy virus |
| HIV-1 | human immunodeficiency virus type 1 |
| HTLV-1 | human T-cell leukemia virus type 1 |
| MLV | murine leukemia virus |
| MPMV | Mason-Pfizer monkey virus |
| RSV | Rous sarcoma virus |
| SIV | simian immunodeficiency virus |
| SNV | spleen necrosis virus |
| TMV | tobacco mosaic virus |
| VSV | vesicular stomatitis virus |
| WDSV | walleye dermal sarcoma virus |

Retroviral Genes and Proteins

| | |
|---|---|
| CA | capsid domain of Gag |
| CTD | C-terminal domain (of CA) |
| CTRS | cytoplasmic targeting and retention signal |
| Env | envelope glycoprotein |
| Gag | structural polyprotein (*g*roup-specific *a*nti*g*en) |
| IN | integrase |
| ISD | internal scaffold domain |
| MA | matrix domain of Gag |
| MBD | membrane binding domain (of MA) |

| | |
|---|---|
| MHR | major homology region (of Gag) |
| NC | nucleocapsid domain of Gag |
| NTD | N-terminal domain (of CA) |
| p2 | minor RSV Gag domain |
| p10 | minor RSV Gag domain |
| p12 | minor MLV Gag domain |
| Pol | enzymatic polyprotein |
| PR | retroviral protease |
| RT | reverse transcriptase |
| SP | minor RSV Gag domain (*sp*acer) |
| Ψ (psi) | genome packaging signal |

Cellular Proteins

| | |
|---|---|
| ABCE1 | ATP-binding cassette protein E1 |
| CREB | cAMP response element-binding protein |
| CRM1 | Chromosome region maintenance 1 (yeast), exportin 1 (human) |
| FKBP | FK506 binding protein |
| GCN4 | General control of amino acid synthesis 4 |
| HP68 | Host protein, 68 kDa |
| SCAN | *S*REZBP, *C*tfin51, *A*W-1 (ZNF174), and *N*umber 18 domain |
| TIP47 | Tail-interacting protein, 47 kDa |

Reagents and Buffers

| | |
|---|---|
| BMH | Bismaleimidohexane |
| DEAE | Diethylaminoethyl-cellulose |
| DMEM | Dulbecco's modified Eagle's medium |

| | |
|---|---|
| DTT | Dithiothreitol |
| EDTA | Ethylenediaminetetraacetic acid |
| MES | 4-Morpholineethanesulfonic acid |
| PBS | Phosphate-buffered saline |
| PCMB | p-chloromercuribenzoic acid |
| PMSF | phenylmethylsulfonyl fluoride |
| RIPA | radioimmunoprecipitation assay buffer |
| SDS | sodium dodecyl sulfate |
| TBST | Tris-buffered saline plus Tween |
| TCEP | Tris(2-carboxyethyl) phosphine |
| Tris | Tris(hydroxymethyl)aminomethane |
| VSB | virion suspension buffer |

Miscellaneous

| | |
|---|---|
| AIDS | acquired immune deficiency syndrome |
| ATP | adenosine triphosphate |
| Cryo-EM | electron cryo-microscopy |
| dsDNA | double-stranded DNA |
| ESCRT | endosomal sorting complex required for transport |
| GFP | green fluorescent protein |
| LMB | leptomycin B |
| MTOC | microtubule organizing center |
| MW | molecular weight |
| MWCO | molecular weight coefficient |
| NES | nuclear export signal |
| NMR | nuclear magnetic resonance |

| | |
|---|---|
| nt | nucleotide |
| PCR | polymerase chain reaction |
| PIC | pre-integration complex |
| $PI_{4,5}P_2$ | phosphatidyl inositol-(4,5)-bisphosphate |
| SDS-PAGE | SDS polyacrylamide gel electrophoresis |
| SEM | scanning electron microscopy |
| ssDNA | single-stranded DNA |
| TEM | transmission electron microscopy |
| VLPs | virus-like particles |
| VSV-G | vesicular stomatitis virus envelope glycoprotein |

# CHAPTER ONE

# INTRODUCTION

## VIRUS STRUCTURE

### Basic properties of virus structure

A virus particle, or virion, is a structure that assembles within an infected cell, exits, and then enters and infects a new cell. As such, the virus particle must meet two conditions. First, the virion must be able to withstand the extracellular environment for some period of time without loss of infectivity. Second, the virion must contain all of the components required for infectivity. These required components may include nucleic acid, protein, and lipid molecules.

The fundamental component of all viruses is a nucleic acid genome that transmits the virus's genetic information. The remaining components exist to protect and support the genome, to assist it to replicate itself, and to move it from cell to cell. The genome may consist of single- or double-stranded DNA or RNA, and single-stranded viral genomic RNA may be either positive- or negative-sense. As nucleases are ubiquitous in the cytoplasm and in the extracellular environment, the genome must be well protected within the virion in order to escape degradation.

All virus particles also include viral proteins. Most of the proteins in a particle will be viral proteins, although many viruses package some host cell proteins as well. Some simple viruses, such as tobacco mosaic virus (TMV), have only a single type of viral protein present in virions, while the prototypical poxvirus vaccinia virus contains at least 63 viral proteins (132). Viral proteins

fall into two broad categories: structural and nonstructural proteins. Structural proteins are those that are necessary to form the virus particle, while nonstructural proteins, such as most viral enzymes, are not required to form the virus particle but are necessary for some other part of the viral life cycle. The division is not entirely simple, as many structural proteins have additional, non-structural functions. The structural proteins form a protective coat surrounding the non-structural proteins and the genome.

Some viruses, known as enveloped viruses, also have an outer lipid envelope derived from the host cell membrane. Acquisition of this membrane is referred to as “budding.” The final envelope may be acquired at the plasma membrane or at internal cell membranes. The viral envelope typically contains viral and cellular membrane proteins. While host glycoproteins may be incorporated at random simply because they are present at budding sites, viral glycoproteins tend to be specifically incorporated into the envelope, and many viruses are actually unable to bud in the absence of their envelope proteins. In addition to protecting the contents, the envelope facilitates entry of the particle into an uninfected cell; it thus serves both structural and non-structural functions.

Virus structure is important at several stages of the viral life cycle. The structure ensures that all necessary components are incorporated into the assembling particle and maintains the stability of the virus particle, protecting the contents from environmental damage after release from the cell. Viral structure is also important after entry into a newly infected cell. In most cases, the particle must disassemble to release its functional components, a process known as uncoating. In more complex cases, such as in the rotavirus family, parts of the virus particle remain assembled and act as a factory for synthesis

of messenger RNA (70) or other viral molecules. In some cases, mutations that enhance virion stability can actually decrease infectivity, presumably by interfering with uncoating (40).The structure must balance these different functions in order to optimize infectivity.

The presence of an envelope also influences virus structure. Since the envelope provides an additional layer of protection and containment, the protective function of the protein coat can be reduced. Non-enveloped viruses thus possess a dense and tightly packed protein coat called a capsid, made up of one or more layers of protein, to prevent degradation of the genome, while most enveloped viruses have a nucleocapsid within the envelope, with a relatively loosely packed protein coating the genome. Virus structure thus depends on the biological requirements of the particular virus.

**Virus structure and symmetry**

The range of available virus structures is constrained by the inherent properties of the proteins that make up the viral coat. Most viral genomes encode only a few coat proteins, protein-protein interactions are very specific, and each protein molecule can make only a limited number of protein-protein contacts. One possible way to overcome these limitations is to arrange the coat protein subunits so that every subunit makes the same contacts with its immediate neighbors as every other subunit of the same type; in other words, to build a structure with extensive symmetry. Crick and Watson originally predicted in 1956 that viruses must have either helical or cubic (i.e. rotational) symmetry (27), and as virus structures were solved both forms of symmetry were observed.

In helical viruses, the molecules of a structural protein interact with the genome and each other to form a single helical structure, the length of which is determined by the length of the nucleic acid molecule enclosed ("open" structure). Helical symmetry is found in both non-enveloped viruses (mostly plant viruses, such as TMV, and bacteriophages) and enveloped viruses (including animal viruses, such as rhabdoviruses).

Cubic symmetry is more complex. Briefly, the requirement that each protein subunit be able to make identical contacts with its neighbors means that the virus must have axes of symmetry such that every subunit is indistinguishable from the others when rotated around those axes. Mathematically, only Platonic solids (regular convex polyhedra) can fulfill this requirement, and the five Platonic solids have only three patterns of symmetry among them. The three cubic symmetries available to viral coats are tetrahedral, cuboidal/octahedral, and dodecahedral/icosahedral. In practice, all regular virus structures solved to date have icosahedral symmetry. An icosahedron is a convex geometric solid with twenty faces (each an equilateral triangle), thirty edges, and twelve vertices. An icosahedral virus thus has six fivefold axes of symmetry (one through each pair of opposite vertices of the icosahedron), ten threefold axes of symmetry (one through the center of each pair of faces), and thirty twofold axes of symmetry (one bisecting each pair of edges). Since proteins are not symmetrical, each face of the icosahedron must consist of three protein subunits with symmetry about the threefold axis through the face. Therefore, in the simplest case, an icosahedral virus must consist of 60 (3 x 20) protein subunits (27). In such a virus, the subunits make pentameric interactions at each vertex, trimeric interactions at the center of each face, and dimeric interactions at each edge. Each subunit makes all

three interactions, and each interaction of a given type is identical for each subunit (**Figure 1.1**).

As there are limits to the size of viral structural proteins, a virus with a large amount of nucleic acid might require more than 60 such proteins to enclose the viral genome. In fact, while the smallest icosahedral viruses, such as parvoviruses, do contain the predicted 60 subunits, other icosahedral viruses are composed of many more. Placement of these extra subunits requires triangulation, or quasi-equivalence, in which some subunits will undergo subtle deformations in order to make inter-subunit contacts that are slightly different from those made by other subunits. Triangulation is modeled by asking how equilateral triangular “facets,” smaller than the faces of an icosahedron but each still consisting of three protein subunits, can be used to enclose a space while maintaining icosahedral symmetry. This is accomplished by building icosahedral fullerene structures, with subunit pentamers at each fivefold axis interspersed with hexamers between the vertices. Each subunit in a hexamer makes a six-fold interaction at the center rather than a fivefold interaction, meaning that the protein-protein contacts, while similar, cannot be identical; the three- and two-fold interactions are also very slightly different depending on whether the subunits involved are in pentamers or hexamers. These structures are categorized by the triangulation number T, where T represents the number of facets occupying each triangular face of the icosahedron. As each facet has three protein subunits, the total number of subunits per particle is 60T. Quasi-equivalence thus expands the range of sizes available to viruses without significantly increasing the required sizes of the viral proteins.

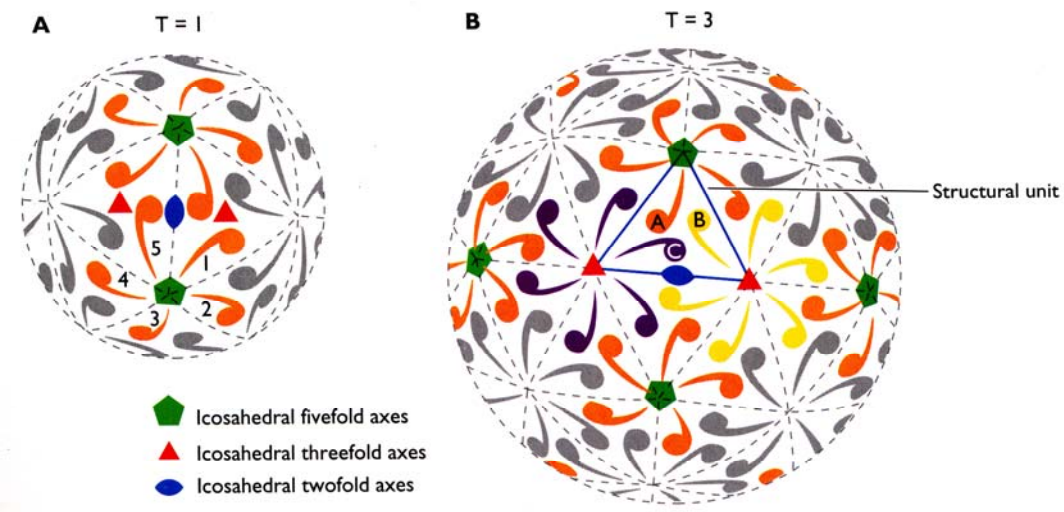



**Figure 1.1**: Icosahedral symmetry. A. T=1 virus with sixty protein subunits (orange commas). Symmetry axes are noted. B. T=3 virus with 180 protein subunits. Pentamers are in orange, hexamers are in yellow and purple. Symmetry axes are noted. Figure take from reference (38).

The sizes of icosahedral viruses are constrained by geometry. Only certain T numbers are compatible with icosahedral symmetry. The allowed numbers are those for which $T=h^2+hk+k^2$, where *h* and *k* are integers representing translations along the 0° and 60° axes of an equilateral triangular lattice (20). Icosahedral viruses employing classical quasi-equivalence may therefore have T numbers of the series 1, 3, 4, 7, 9…; T numbers up to 219 (h=7, k=10) have been documented (131). The number of subunit pentamers is fixed at 12 (20), whereas the number of hexamers will increase with the T number; therefore, while a T=1 virus is composed only of pentamers, even a T=3 virus will have twice as many subunits in hexamers as in pentamers (**Figure 1.1**), while in large icosahedral viruses the pentamers will be vastly outnumbered. Therefore, large icosahedral capsids may be described as a lattice of hexamers with symmetrically placed pentamers.

**Exceptions to icosahedral symmetry**

While many virus structures conform to the rules above, viruses may also evade these principles in various ways by employing non-standard bonding. For example, the innermost protein shell of some reoviruses, while icosahedral, has a "disallowed" triangulation number of 2. This is accomplished by "geometrical quasi-equivalence," in which the inner coat protein VP3 is able to make two somewhat different sets of contacts and thus occupy two different structural positions (48). Other viruses have "pseudo-symmetry;" the classic example is the picornavirus family. Picornaviruses, such as poliovirus, contain 60 structural units, like a T=1 virus, but each unit is made up of four different proteins, three on the surface (VP1, VP2, VP3) and

one underneath (VP4). Because the three surface proteins, having different amino acid sequences, cannot make strictly quasi-equivalent bonds, and because of the hidden VP4 protein, picornaviruses are said to have pseudo-T=3 symmetry (38). Certain bacteriophages, such as T4 phage, have elongated icosahedral heads with different T numbers for the end caps and the midpiece; for T4 phage, $T_{end}$=13 and $T_{mid}$=21 (39). Icosahedral symmetry or pseudo-symmetry is thus employed by a number of viruses that do not form strictly icosahedral capsids.

Mathematically, a flat hexagonal lattice can be bent into a closed surface by the introduction of twelve pentameric “defects” at various positions without regard to icosahedral symmetry; therefore, by stretching the principle of quasi-equivalence, a virus could employ pentamers and hexamers of the same subunit protein(s) to form coats of various shapes (**Figure 1.2**) (47). It is thus possible for viruses to build many closed structures without strictly observing the principle of quasi-equivalence.

In addition to the variations on helical and icosahedral symmetry of viral coats, many viruses employ additional structural elements. The innermost coat of the bluetongue particle has been mentioned, but in addition to the T=2 shell of VP3, bluetongue cores have an outer layer of VP7 protein in a T=13 lattice, with a pentameric λ2 spike at each vertex (48). This core is further enclosed by an outer capsid layer consisting of three additional proteins (34). Enveloped viruses may, in addition to a protein-studded lipid envelope and a nucleocapsid, contain an intermediate matrix protein layer lining the inner surface of the envelope; influenza (orthomyxovirus) M1 protein is an example. The matrix protein may interact with the cytoplasmic tails of the membrane proteins, with the lipid bilayer itself, or with both (as does M1) (38). It is


**Figure 1.2**: Characteristic core shapes of three different retroviruses modeled as fullerene structures with a hexagonal lattice closed by 12 pentagonal “defects.” The placement of the pentagons determines the core shape. Figure taken from reference (72).

frequently responsible for the shape of the virus particle and may be sufficient to promote budding of virus-like particles of characteristic size and shape even in the absence of a nucleocapsid, as is the case for Ebola virus (55). These additional elements add further layers of complexity to virus structures.

## RETROVIRUSES

### Overview of retroviral biology

Retroviruses are a large and diverse family of animal viruses that cause significant disease in humans and in a number of domestic animal species. The best known retrovirus is Human Immunodeficiency Virus Type 1 (HIV-1), the causative agent of Acquired Immune Deficiency Syndrome (AIDS); as of 2007, approximately 33 million people worldwide were thought to be infected, with 2.5 million new infections and 2.1 million AIDS-related deaths in the previous year (128). HIV-1, like most retroviral infections, is chronic and incurable, but the progression to AIDS and death can be quite slow; these properties have generated immense interest in the development of anti-retroviral therapies, some of which have proven highly effective. However, the rapid development of drug resistance combined with the requirement for lifelong treatment requires continuous development of new and safer anti-viral drugs. Improved understanding of various aspects of retroviral biology will facilitate the development of improved therapies.

The hallmark of retroviral biology is reverse transcription, in which the virus violates the previously-held “central dogma of molecular biology” (26) by synthesizing DNA from an RNA template. This is accomplished by the viral

enzyme reverse transcriptase (RT), which makes a DNA copy of the +-sense single-stranded RNA genome during the early stages of infection. Reverse transcription is shared by the closely related spumaviruses, or foamy viruses (e.g. Human Foamy Virus), which contain a linear DNA genome generated by reverse transcription in the producing host cell, and by the more distantly related hepadnaviruses (e.g. Hepatitis B virus), whose gapped circular DNA genome is also made by reverse transcription (although the details of the RT mechanism are somewhat different) (8). Because RT lacks the 3' exonuclease "proof-reading" domain found in most cellular DNA polymerases, it has a high error rate (approximately $1 \times 10^{-4}$ per base pair) (119). This lack of fidelity has obvious drawbacks – each replication cycle produces many defective virions – but allows the virus to evolve rapidly to evade the host immune response and develop drug resistance.

The family *Retroviridae* is currently divided, based on evolutionary relationships, into seven genera: the alpharetroviruses (typified by Rous Sarcoma Virus, RSV), betaretroviruses (Mason-Pfizer Monkey Virus, MPMV), gammaretroviruses (Murine Leukemia Virus, MLV), deltaretroviruses (Human T-cell Leukemia Virus type 1, HTLV-1), epsilonretroviruses (Walleye Dermal Sarcoma Virus, WDSV), lentiviruses (HIV-1), and spumaviruses (Human Foamy Virus, HFV). The spumaviruses are most distantly related to the others and are classified as a separate subfamily, the spumaretroviridae, while the other six genera make up the subfamily orthoretroviridae (37).

All retroviruses share three genes: *gag*, *pol*, and *env*. The *gag* gene encodes the major structural protein Gag (*g*roup-specific *a*nti*g*en), the *pol* gene encodes the viral enzymes RT and integrase (IN), and the *env* gene encodes the viral glycoprotein Env. Either *gag* or *pol* also encodes the viral

protease PR. The *gag* and *pol* gene products are typically translated from the same open reading frame of a full-length unspliced viral RNA, such that both Gag and Gag-Pol are produced by either a frameshift (e.g. RSV) or a read-through (e.g. MLV) mechanism. Both Gag and Gag-Pol are polyproteins that are cleaved post-translationally by the viral protease PR. Env is translated from a separate spliced mRNA. The "simple" retroviruses have only these three genes, while the "complex" retroviruses have additional "accessory" genes that are beyond the scope of this introduction (**Figure 1.3**). The betaretroviruses, deltaretroviruses, epsilonretroviruses, lentiviruses, and spumaviruses have complex genomes, while the alpharetroviruses and gammaretroviruses have simple genomes. The lentiviruses are further distinguished by their ability to infect non-dividing cells (which does not wholly depend on the accessory gene products (12)), whereas most of the other orthoretroviridae infect only dividing cells. Lentiviruses also tend to cause immunodeficiency disease, while the other orthoretroviruses cause mostly neoplastic disease.

**The retroviral life cycle**

The retroviral life cycle is divided into two stages, early and late (**Figure 1.4**). The early stage includes those steps which lead to infection of a host cell. It begins when the incoming virus particle binds to one or more cellular receptors via the Env protein, which then undergoes a conformational change to initiate fusion of the viral envelope with the host cell membrane. Fusion may occur at the plasma membrane (e.g. HIV-1) or following receptor-mediated endocytosis and endosomal acidification (e.g. RSV (89)). The viral core is released into the cytoplasm, where reverse transcription takes place. The core

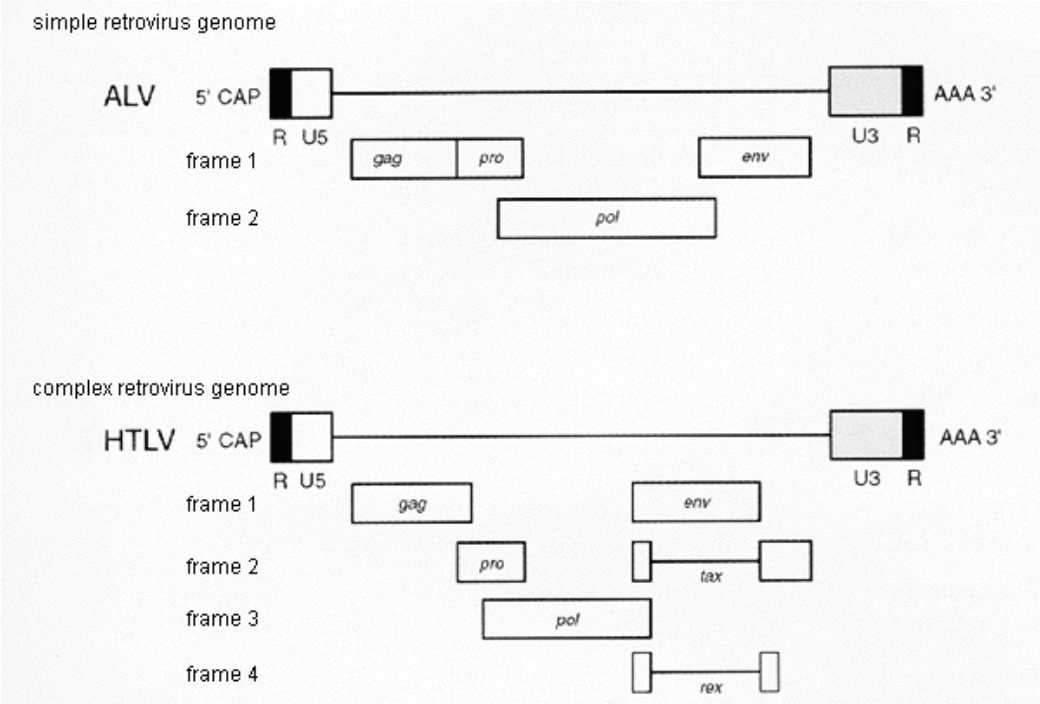



**Figure 1.3**: Comparison of simple (ALV aka RSV) and complex (HTLV) retroviral genomes, Vertical spacing represents different open reading frames, horizontal gaps represent open reading frames accessed by alternative splicing. Figure taken from reference (121).

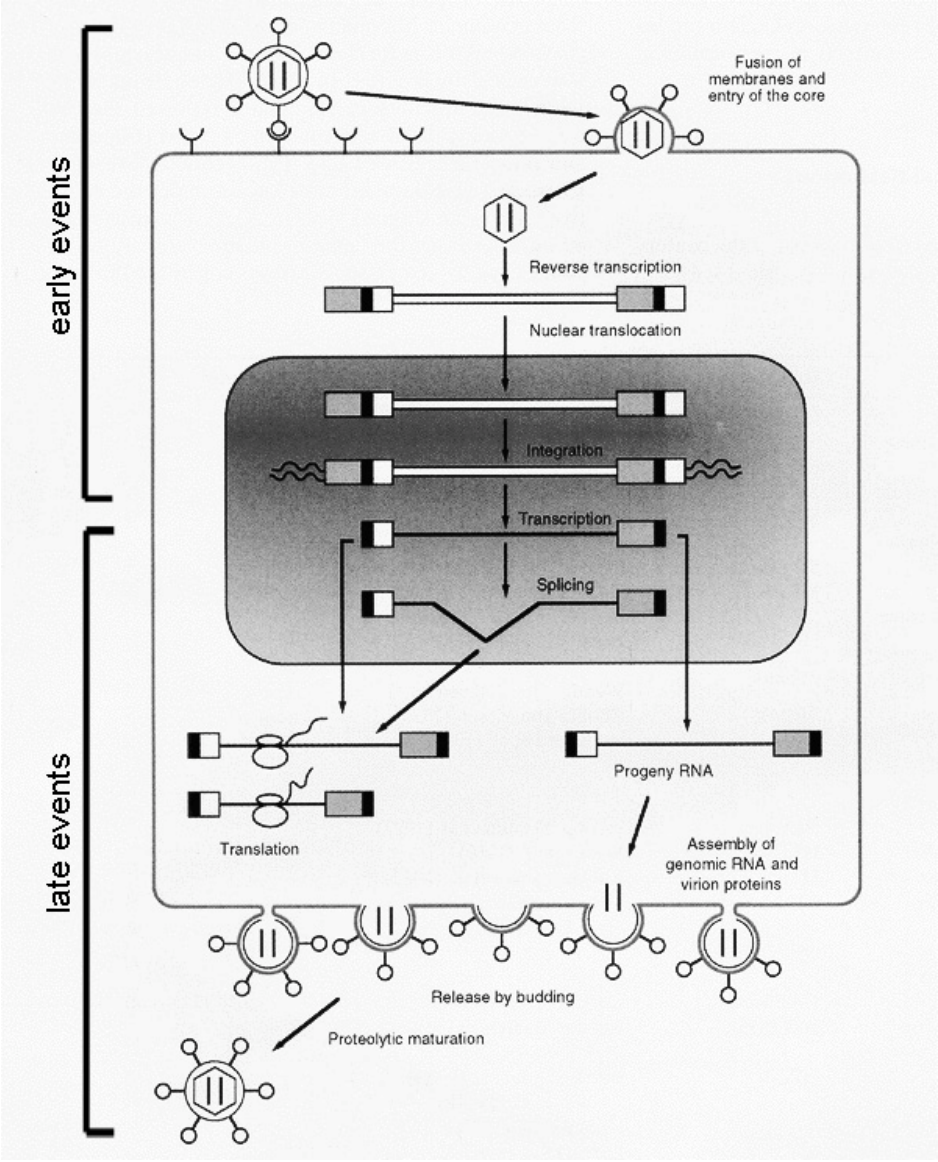



**Figure 1.4**: Schematic representation of retroviral life cycle showing early and late events of replication. Figure adapted from reference (121).

then “uncoats” to release the “pre-integration complex” (PIC), which consists of the viral DNA, integrase, and other viral and cellular components. The PIC enters the nucleus by active importation (lentiviruses) or during mitosis (simple retroviruses). There, the viral DNA is integrated into the host’s chromosomal DNA. Integration marks the end of the early stage of replication, and cells that contain integrated viral genomes are regarded as infected.

The late stages of retroviral replication are those which produce the virus particle. First, the viral DNA is transcribed by cellular RNA polymerase II, producing both full-length (Gag and Gag-Pol) and spliced (Env and accessory proteins) mRNAs. These mRNAs are exported from the nucleus (full-length mRNAs depend on either special cis-acting sequences or accessory gene products for export) and translated by either free (full-length) or membrane-bound (Env) ribosomes to yield viral proteins. The Gag protein recognizes and binds the full-length viral RNA via a secondary structure-rich 5’ “packaging signal” known as Ψ (psi). The full-length viral RNA serves as both mRNA for Gag and Gag-Pol and as genomic RNA, although it is not clear whether the same molecules are both translated and packaged. The Gag-RNA complex is transported to the plasma membrane by an unknown mechanism. At the membrane Gag assembles into virus particles and also interacts with components of the cellular ESCRT (Endosomal Sorting Complex Required for Transport) complex (125). In uninfected cells, the ESCRT complex is involved in the budding of vesicles into late endosomes (multi-vesicular bodies) and in cytokinesis (86), processes topologically similar to viral budding from the plasma membrane. Retroviral Gag proteins contain “late domains,” generally tetrapeptide motifs, which interact with various ESCRT proteins and recruit them to sites of viral budding; in the absence of viral late domains or of

ESCRT components, budding virus particles display a “late domain phenotype” in which the particle remains tethered to the cell surface by a membrane stalk (**Figure 1.5**).

During budding the Env protein is incorporated into the viral envelope. While envelope protein incorporation can be passive (as when pseudotyping retroviral particles with the VSV-G protein), there is evidence that retroviral budding sites actively recruit Env in a somewhat virus species-specific manner; for example, HIV-1 budding sites recruit both HIV-1 Env and VSV-G proteins, but not RSV Env protein (Marc Johnson, personal communication). HIV-1 may exploit the cellular protein TIP47 in order to recruit HIV-1 Env (75). During or shortly after release, the viral protease is activated and cleaves Gag and Pol polyproteins into their constituent domains. Cleavage by PR results in morphological changes to the particle termed maturation; blocking PR cleavage produces non-infectious “immature” particles. Maturation is the final step of the late stage of retroviral replication.

**The retroviral Gag protein**

The structural protein Gag is both necessary and sufficient for retroviral assembly, meaning that the Gag protein alone, when expressed in cells, will assemble and bud to produce virus-like particles. Such particles lack the viral genome, enzymes, and envelope protein and are thus completely non-infectious, but they resemble authentic immature particles in shape, size, appearance by transmission electron microscopy (TEM), and radial density profile (41). (In the case of RSV, which has PR as part of Gag instead of Gag-Pol, the virus-like particles undergo maturation as well.) Therefore, in

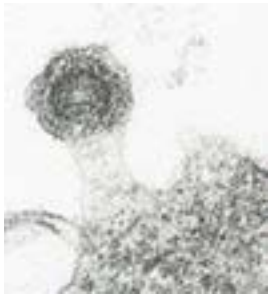


**Figure 1.5**: Budding particle displaying “late domain phenotype,” in which the particle appears to be fully assembled but remains tethered to the cell surface by a membrane stalk. Figure take from reference (116).

discussions of retroviral structure and assembly, the Gag protein is frequently considered alone.

Gag is a polyprotein that is cleaved by the viral protease PR during or shortly after virus budding and release. All retroviruses have three PR cleavage domains in common: the protein-lipid interaction domain MA, the protein-protein interaction domain CA, and the protein-nucleic acid domain NC. The remaining cleavage products vary among retroviral families. High-resolution structures have been obtained for each of the major cleavage products, although not for the entire protein. While there is little sequence identity among retroviral Gag proteins, the secondary and tertiary structures of the major cleavage domains are remarkably similar, allowing comparisons between different retroviruses. Based on these domain structures, Gag is a predominantly helical protein, with no significant amount of β-sheet structure. There is some evidence for intramolecular inter-domain interactions in the soluble state (29), but as the significance of these interactions is not yet clear it remains profitable to consider the cleavage products as independent protein domains.

The MA domain is primarily responsible for interacting with cellular lipids. The retroviral MA domains studied to date have an N-terminal helical "membrane binding domain" (MBD) and a C-terminal extension which is frequently dispensable for assembly and budding (90, 135), and for replication in RSV (92), but not for early steps of replication in HIV-1 (135). In HIV-1, the C-terminus appears to be important for Env incorporation (136), although the site of interaction between MA and TIP47 is in the MBD (75). The MBD, or "M domain" (95), consists of four helices and a short mixed β-sheet packed into a globular domain. A fifth helix connecting to the C-terminus projects away from

the globular domain, and the C-terminus itself is unstructured. The MA-lipid interaction is predominantly electrostatic, with a basic patch on the MBD interacting with negatively charged phospholipids. In addition, the MA proteins of most retroviruses are myristylated at the N-terminus, although those of RSV and equine infectious anemia virus (EIAV) are not. While the fatty acid modification contributes to membrane affinity, the electrostatic interaction predominates (28). In solution structures, unmyristylated HIV-1 MA proteins are monomeric, while in crystal structures the unmyristylated protein is trimeric; the myristylated protein exists in both states in solution, with the myristate sequestered in the monomer form and exposed in the trimer (117). This “myristyl switch” is thought to couple membrane binding to assembly, although it is not clear that MA forms trimers in either the immature or the mature virus particle. HIV-1 and EIAV MA proteins also contain a phosphatidylinositol-4,5-bisphosphate ($PI_{4,5}P_2$) binding pocket that is thought to target assembly to the plasma membrane, and $PI_{4,5}P_2$ binding may also activate the myristyl switch (21, 103).

The CA domain is the major protein-protein interaction domain of Gag. The CA domain of Gag drives both mature and immature assembly. This is obvious in the case of mature assembly, in which the outer shell of the core consists of CA alone. For immature assembly, the CA domain of chimeric Gag molecules is what determines Gag-Gag interactions. Chimeric Gag can co-assemble with and complement other Gag molecules with homologous CA domains, but not those with non-homologous CA domains (2, 71). The CA domain also determines the size and core morphology of immature virus-like particles (2).

The CA protein consists of independently folded N- and C-terminal domains (NTD and CTD) connected by a short flexible linker (19, 62) (**Figure 1.6**). The N-terminal domain consists of a short N-terminal β-hairpin and seven α-helices. The β-hairpin is stabilized by a salt bridge between the amino terminus of the conserved initial proline residue of CA and the side chain of a conserved aspartate in helix 3. Therefore, the β-hairpin forms only after PR cleavage (which liberates the amino terminus of proline 1) and does not exist in the Gag precursor protein.

Several studies have addressed the NTD structure prior to PR cleavage. An NMR structure has been obtained for an HIV-1 Gag fragment containing MA and the N-terminal domain of CA (118). The β-hairpin region of the NTD (**Figure 1.6**, arrows) is unfolded, as predicted when the N-terminus of proline 1 is not free to form the anchoring salt bridge. This unfolding produces an approximately 2 Å shift of helix 6 and the helix 4-5 loop, but the remainder of the NTD structure is similar to that obtained for the mature NTD. A solution structure was recently obtained for a similarly “immature” MLV construct containing the NTD of CA and the upstream domain p12 (67). Like the HIV-1 structure, the NTD portion of the protein resembles the structure of the mature NTD with the β-hairpin unfolded. The upstream p12 domain is unstructured, highly mobile, and does not interact with the NTD. This is in striking contrast to a structure of the RSV NTD with an N-terminal 25 amino acid extension into the upstream p10 domain. The RSV extended NTD crystallized as a dimer with an extensive intermolecular p10-CA interface (91) (**Figure 1.7**). The significance of this dimer was not certain prior to the work described in this thesis, but it seemed unlikely that the RSV extended NTD dimer could be representative of the state of the NTD in immature virus particles. Thus, while

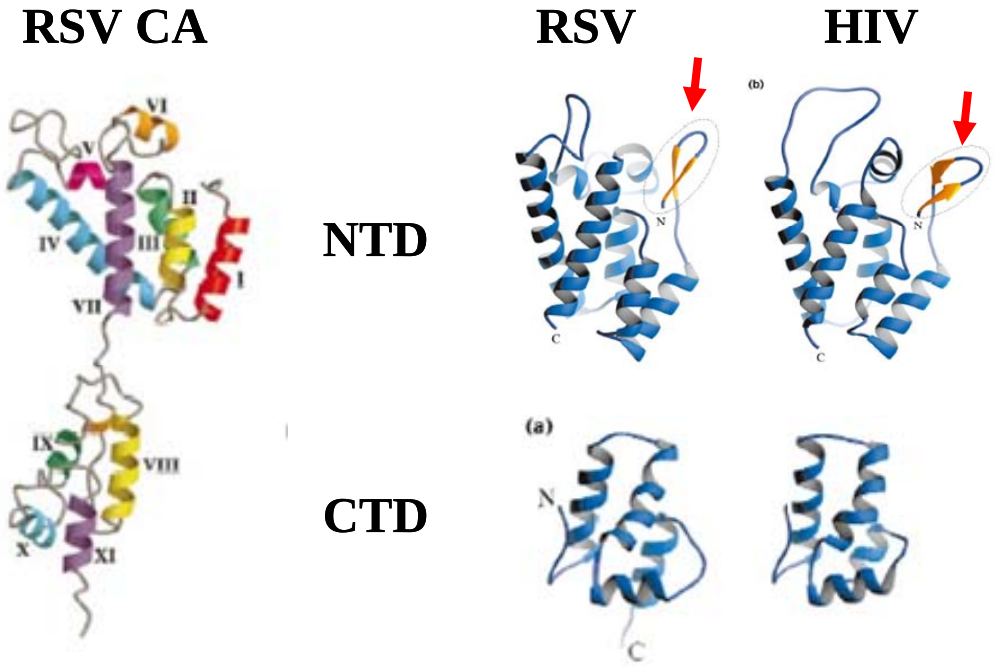



**Figure 1.6**: **Left**: Structure of full-length RSV CA protein. **Right**: Comparison of NTD and CTD from RSV and HIV. Note similarities of tertiary structure despite lack of sequence homology. Red arrows indicate N-terminal β-hairpins. Figure adapted from reference (64).

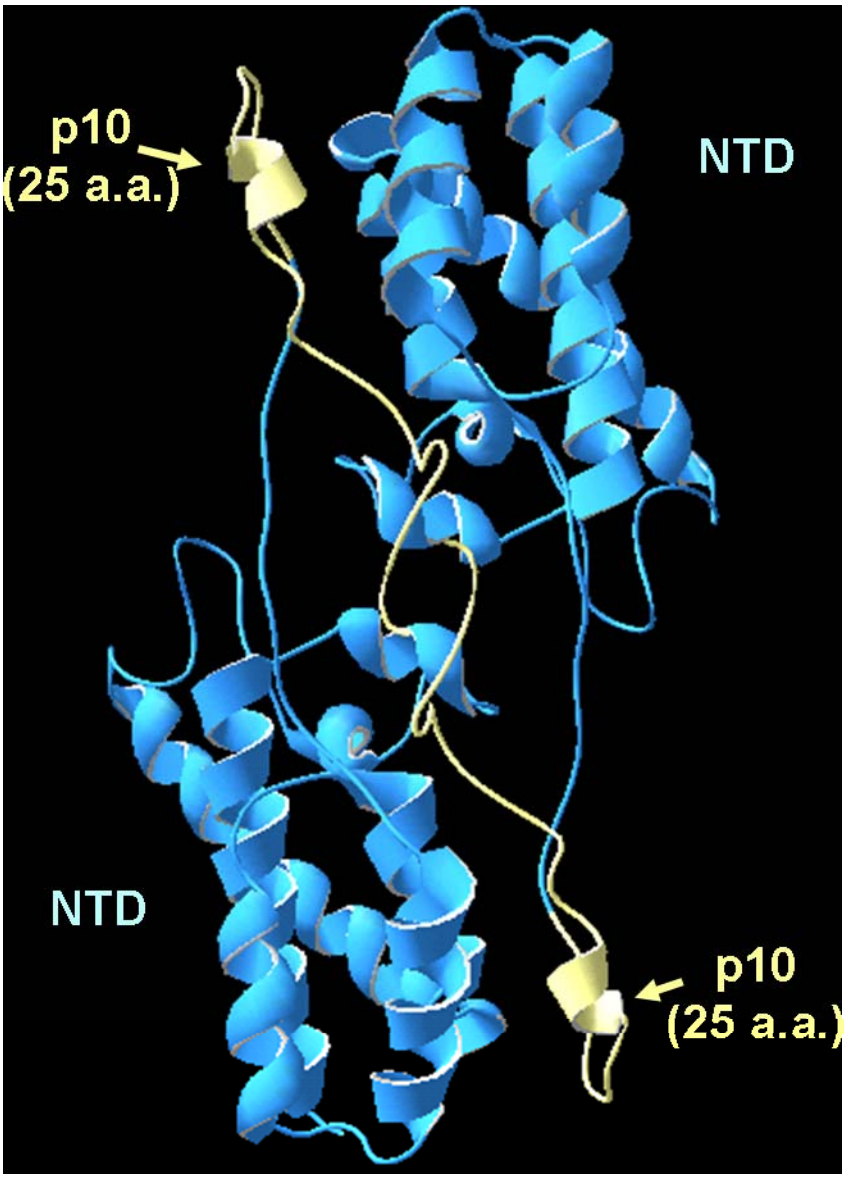



**Figure 1.7**: RSV extended NTD structure (91). NTD residues are in blue and p10 residues in yellow. Note the lack of β-hairpin and the large p10-CA interface.

it is certain that the conserved β-hairpin is found only in the post-cleavage form of the NTD, the significance of hairpin formation is not clear.

In contrast to the situation for the NTD in the context of Gag, the structure of the mature NTD in the context of the mature virus particle has recently been established. The NTD of murine leukemia virus (MLV) has been crystallized as a hexamer that is believed to represent its state in the mature core (87, 88). The hexamer is formed by interactions among the first three helices of each subunit, as previously predicted by mutagenesis of the HIV-1 NTD (124). Interactions among the β-hairpins may contribute to the hexamer interface. The packing of the hexamers within the crystal may also be relevant to assembly (87) (S.P. Goff, personal communication), although this remains in dispute. Both the structure and the role of the NTD in mature particles are thus well characterized.

The C-terminal domain of CA consists of four helices. The CTDs of HIV-1 and EIAV, but not that of HTLV-1, crystallize as dimers mediated by hydrophobic interactions between side chains on helix 2. The RSV CA CTD has been crystallized as a monomer (64) and as a dimer (Richard Kingston, personal communication). HIV-1 CA is also dimeric in solution (42) due to the CTD-CTD interaction, although EIAV CA is monomeric in solution (7) and dimerizes only in the crystal structure. Most retroviral CA proteins (although not that of RSV) have a pair of cysteine residues in helices 3 and 4 of the CTD that form an intramolecular disulfide bond in the HIV-1 and EIAV crystal structures. The disulfide bond is not present in freshly prepared HIV-1 particles (81), in the monomeric soluble EIAV CA (7), in the solution structure of HTLV-1 CA (62), or in the recently obtained NMR structure of a monomeric mutant of

the HIV-1 CTD (129). The correlation between disulfide formation and dimerization is intriguing, but its significance remains debatable.

The CTD also contains the major homology region (MHR), a 20 amino acid sequence with unusually strong sequence identity among the retroviruses (and, to a lesser extent, in some retrotransposons (24, 93, 100) and in the distantly related hepadnaviruses (141)). The MHR occupies the N-terminal loop and part of the first helix of the CTD. Mutation of the MHR can be rescued by second-site mutations in both the NTD and CTD (98), suggesting a complex role for the MHR in assembly and infectivity. Recently, a mammalian SCAN domain structure was found to have significant structural homology to the retroviral CTD (53). The SCAN dimer appears to be a “domain-swapped” analog of the CTD, in which helix 1 is “exchanged” between the subunits and makes dimer contacts very similar to the intramolecular contacts in the monomer. The HIV-1 CTD can be manipulated by deletion of an amino acid in the CTD helix 1-2 loop to crystallize as a similar domain-swapped dimer (54), and MHR residues are involved in the interface. While the significance of this finding is not clear, it is an intriguing model as it assigns an assembly function to the MHR.

The NC domain of Gag is relatively short and unstructured. The domain has a strong net positive charge, and the NC-RNA interaction, like that between MA and membrane phospholipids, is thought to be largely electrostatic. NC domains of all but the spumaviruses also contain one or two zinc finger structures (sometimes called “zinc knuckles” due to their length, which is short compared to canonical zinc finger domains), consisting of a CCHC motif coordinating a zinc (II) ion. The zinc fingers are required for specific packaging of viral RNA during assembly (122) and also for proper

reverse transcription and integration in the early stages of the retroviral life cycle . The remainder of the NC protein is unstructured in its unbound state; additional folding is apparently induced by binding to RNA. NMR structures of psi sequences bound to NC show the zinc fingers interacting with various RNA secondary structural elements (6, 30, 138). NC has occasionally been called dispensable for budding (94), although most investigators find that it is required for particle production (108, 109, 126), most likely because NC-RNA binding initiates Gag assembly. Recent studies of the RSV Gag protein suggest that the first four amino acids of RSV NC are part of a structural element that also includes the SP linker domain that joins the CA and NC domains and the C-terminal tail of CA (60). Therefore, NC may have as many as three separate functions in retroviral particle production.

**Special aspects of RSV Gag**

As this laboratory uses RSV as a model to study retroviral assembly, some special consideration of the RSV Gag protein is warranted. In addition to the three primary domains (MA, CA, and NC), RSV contains four additional PR cleavage domains: p2 and p10 between MA and CA, SP between CA and NC, and PR itself C-terminal to NC (**Figure 1.8**). The p2 domain has no known structural significance but contains the primary RSV late domain, the Nedd4 ubiquitin ligase-binding motif PPPY (63). The p10 domain has at least two functional elements. The N-terminal half of p10 contains the secondary late domain, the Alix-binding motif LYPSL (Kari Dilley, personal communication). The last 25 amino acids of p10 are required for correct RSV particle morphology (56, 58) and are those included in the extended NTD structure previously described; these 25 residues are the focus of Chapter Three of this

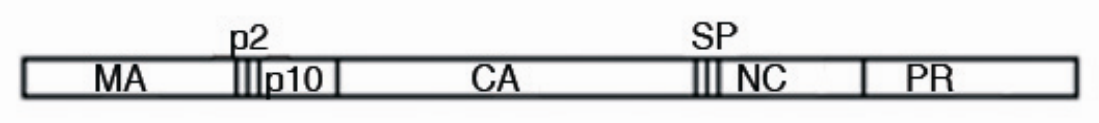



**Figure 1.8**: The RSV Gag protein. Vertical lines represent PR cleavage sites. Note that the 22 amino acid p2 and the 12 amino acid SP domains undergo an additional internal cleavage.

thesis. The SP domain has not been well characterized, but it too appears to be required for spherical particle morphology: interruption of a putative structural element consisting of the unstructured C-terminal tail of CA, SP, and the N-terminal four amino acids of NC has been shown to produce cylindrical particles when expressed in cells (60). Finally, for RSV, the PR coding sequence is before the Gag-Pol frameshift, so PR is part of Gag rather than of Gag-Pol. This means that an RSV particle contains many more molecules of PR than most retroviral particles, but this has no known significance for the function of Gag itself.

**Retroviral particle structure**

The structural biology of retroviruses is complicated by the fact that retroviral particles exist in two forms: mature and immature (**Figure 1.9**). Maturation is a consequence of cleavage by the viral protease PR, which separates the domains of Gag during or shortly after viral budding. The mature particle is the infectious form, and immature particles are rarely observed under natural conditions, although they may be produced by disrupting the activity of the retroviral PR protein by mutation or drug treatment. Little is known about the timing of PR activation or maturation. The infrequent observation of immature particles suggests that maturation must occur almost simultaneously with budding. However, the tethered particles characteristic of the “late domain phenotype” appear to be immature even when an active protease is present, suggesting that maturation cannot occur until budding is complete. This delay suggests that the immature particle, while not infectious, is a necessary intermediate worthy of study.

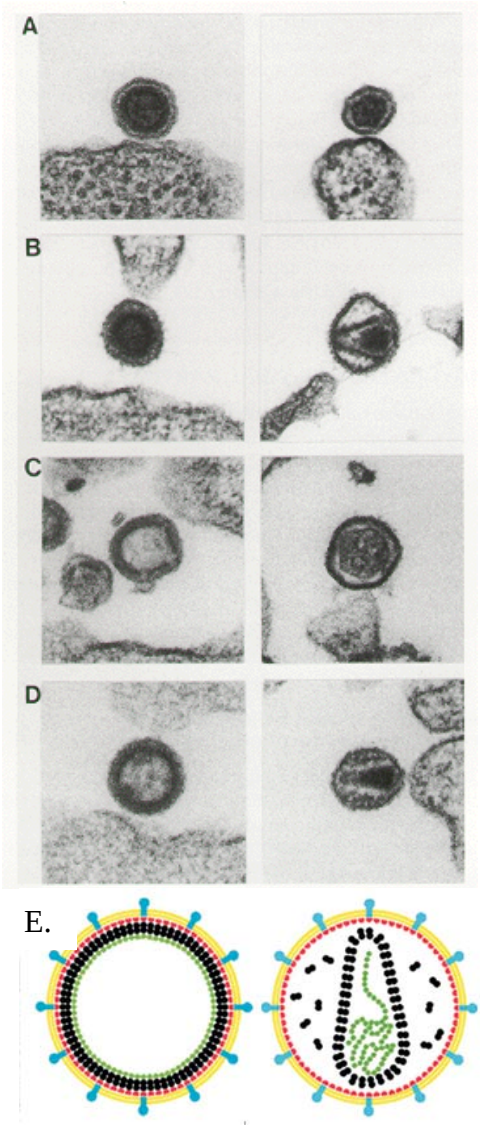



**Figure 1.9**: Morphological maturation of retroviral particles. Immature form is on the left, mature form is on the right. **A.** MLV, **B.** MPMV, **C.** HTLV-1, **D.** HIV. EM images taken from reference (116). **E.** HIV maturation cartoon. Immature particle is on the left, mature particle is on the right. Red dots represent MA, double black dots represent CA, green dots represent NC. Diagram taken from reference (10).

Morphological maturation is dramatic at the level of electron microscopy. Viewed by thin-section transmission electron microscopy (TEM), the mature particle has a central, electron-dense core, which consists of a shell of CA protein enclosing the dimeric RNA genome and associated NC protein. Env protein spikes are clearly visible projecting from the outer surface of the envelope, and the MA protein coats the inner surface of the lipid envelope. The arrangement of the MA molecules remains to be elucidated; while the isolated protein forms trimers in crystal structures and hexamers in a 2D lattice (5), no evidence yet exists for an organized MA lattice in retroviral particles (11). The CA shell has been studied extensively and is thought to be based on a "fullerene" model in which a hexameric lattice is closed by twelve pentameric defects (47) (**Figure 1.2**). Numerous reconstructions of two-dimensional crystals and cylindrical CA assemblies suggest that the NTD of CA forms hexamers that are linked together by CTD dimers (46, 47, 72, 80) (**Figure 1.10**). The hexameric crystal structure of the NTD of MLV CA (88) supports the reconstructions and is believed to represent the mature NTD hexamer. The CTD dimer in 2-D reconstructions of HIV-1 CA lattices does not appear to be a perfect match for either the canonical "face-to-face" dimer or the domain-swapped crystallographic dimer, although it is closest to the former (44). The structure of the shell's contents – the RNA, and the NC, RT, and IN proteins – remains largely unexplored.

In contrast to mature particles, immature particles examined by thin-section TEM have an electron lucent center and an electron-dense ring underlying the lipid envelope. The identity of the dense ring is not clear; it may represent either the closely packed CA domains of Gag or the associated NC domains and RNA. The Gag molecules are thought to be arrayed radially, with

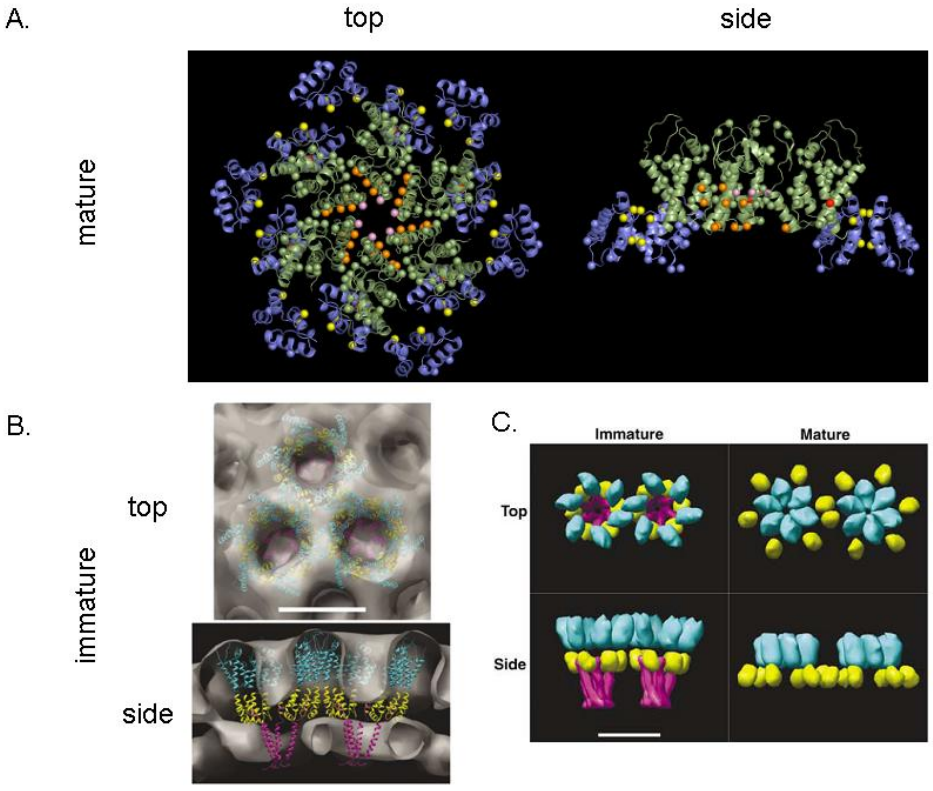



**Figure 1.10**: **A.** HIV CA reconstruction representative of mature particle. NTD hexamer is in green with interface depicted by red dots. CTD dimers are in blue with interface depicted by yellow dots. Figure taken from reference (44). **B.** Immature HIV particle cryotomography with Gag modeled in. NTD hexamers are blue, CTD dimers are yellow, downstream "pillar" structures are magenta. **C.** Comparison of immature and mature lattice arrangements. Color scheme as in **B**. Panels **B** and **C** taken from reference (130).

the MA domain contacting the lipid envelope and the NC domain interacting with the viral RNA toward the center of the particle. Radial density profiles of immature particles obtained by cryo-electron microscopy support this hypothesis, showing concentric layers of density interpreted as MA, the N- and C-terminal domains of CA, and NC (10, 41, 50, 133). The MA domain layer of the immature particle, as examined by cryo-electron microscopy diffraction (10) and single-particle cryotomography (130), does not display a regular protein lattice. The CA layer is superficially similar to that of the mature particle, with NTD hexamers linked by CTD dimers (130) (**Figure 1.10**). There are, however, significant differences. The immature core contains more molecules of Gag than the mature CA shell does of CA. The exact numbers remain controversial, but the difference may be as much as two-fold for HIV-1 (10, 11). The lattice spacing (the distance between the centers of the hexamers) of the immature CA is correspondingly smaller, 8 nm for immature particles vs. 9 nm for HIV-1 mature particles (10, 11, 47, 72, 130). These differences have two implications. First, since the lattice spacing is different, the CA-CA contacts made by Gag are somewhat different than those made by mature CA. Second, since the number of molecules is different, the immature lattice must at least partially disassemble before the mature lattice forms.

Single-particle cryotomography revealed three surprising features of the immature HIV-1 particle (130). First, the Gag lattice appears to be incomplete, with substantial areas containing no organized Gag at all; this property may account for the diverging estimates of the number of Gag molecules in the core, as some methods have assumed a complete lattice. Second, the NTD hexamers were larger than expected. Since the immature lattice spacing is smaller than that of the mature lattice, the increased hexamer diameter means

that the immature hexamers must be much more tightly packed than the mature hexamers, presumably due to differences in the CTD-CTD interactions. The hexamers themselves were also surprising: rather than resembling solid disks, as in the hexameric MLV NTD crystal structure, they appeared as rings with large central cavities (**Figure 1.10**). No direct NTD-NTD interactions were observed within the hexamers, although the resolution of this technique is not sufficient to say that none exist. Third, the “spacer” region between CA and NC formed prominent pillar-like structures which lay below the central cavities of the NTD hexamers. As the spacer region of HIV-1 Gag has been reported to form a long α-helix under some conditions(83), it has been suggested that the pillar formations are six-helix bundles, although the resolution was not sufficient to establish this definitively. Whatever their structure, these pillars effectively tether the C-terminal domains of Gag below the NTD hexamers and may thus be responsible for the tight packing of the immature lattice. These three features have also been observed for RSV particles (John Briggs, personal communication) and may thus reflect general principles of retroviral immature structure.

The mechanism by which PR cleavage leads to morphological maturation is still unclear, although improvements in the understanding of the structures of immature and mature particles are beginning to elucidate this process. For example, cleavage between CA and the downstream spacer domain, which separates the CTD from the putative six-helix bundles, presumably allows a rearrangement of the CTD dimers relative to the NTD hexamers. At the N-terminal end of CA, the formation of the β-hairpin after PR cleavage is likely to play a role in maturation. The β-hairpin appears to be critical for mature core assembly. It forms a large part of the NTD-NTD

interface in the MLV hexamer (88), and the salt bridge is required for correct mature morphology for both HIV-1 (123) and Mason-Pfizer Monkey Virus (MPMV) (120) although apparently not for HTLV-1 (9). There is thus a putative candidate at each end of the CA protein for the switch to mature morphology.

For RSV, there is a more explicit link between maturation and PR cleavage at the N-terminus of CA: a sequence upstream of CA is required for immature morphology. Removal of the last 25 amino acids of the p10 domain of Gag results in the formation of cylindrical particles believed to represent mature cores (18) (as purified HIV-1 CA also assembles into tubes *in vitro* (36, 45)). Therefore, the upstream sequence, which would be removed by PR cleavage during the normal life cycle, is directly responsible for the change in morphology. A structure obtained for an RSV NTD extended N-terminally to include the C-terminal 25 amino acids of p10 showed an antiparallel dimer with an extensive p10-CA interface (91) (**Figure 1.7**). As will be shown in this thesis, this structure is in fact relevant to immature assembly.

## RETROVIRAL ASSEMBLY

### Assembly in cells

The process by which retroviral particles assemble in cells is poorly understood. Two general assembly pathways have been described. Most retroviruses appear to assemble in association with the cellular membranes from which they bud. The exact location of assembly and budding is somewhat controversial, particularly for HIV-1 in monocyte-derived cells (99, 127), but in general most retroviruses are thought to assemble at and bud from the plasma membrane in most cell types. In the alternate pathway,

formerly known as type-B or type-D assembly, immature particles assemble in the cytoplasm near the microtubule organizing center (MTOC) and are subsequently transported to the plasma membrane for budding. This pathway is used by the betaretrovirus and spumavirus families. Pericentriolar assembly is dictated by a “cytoplasmic retention targeting signal” (CTRS) located within the Gag protein (22, 134). For MPMV, mutation of the CTRS redirects assembly to the plasma membrane (114), while for primate foamy virus mutation of the CTRS abolishes assembly (35). Re-targeting Gag to the plasma membrane can even rescue assembly-defective MPMV (107). Therefore, assembly at the plasma membrane appears to be the “default” retroviral assembly pathway.

To date, most work on the retroviral assembly process (as opposed to specific requirements for assembly and/or budding) has focused on identification of assembly intermediates. Under some circumstances, assembly within cells tolerates the replacement of the NC domain of Gag with a coiled-coil dimerization domain such as the leucine zipper domain from CREB or GCN4, suggesting that Gag dimerization is sufficient to initiate assembly (1, 56, 137). The particles produced do not package viral RNA (137) and may contain no RNA at all (Alan Rein, personal communication). HIV-1 Gag-containing complexes of discrete sizes have been identified by sedimentation analysis, suggesting an ordered assembly pathway in which the size of the Gag complex increases in a stepwise fashion (32). This process appears to involve the cellular ATP- and RNA-binding protein ABCE1 (33), which interacts with Gag via basic residues in the NC domain (73). This interaction is independent of RNase treatment, although an RNA-mediated interaction is difficult to exclude completely. ABCE1 is recruited to HIV-1

budding sites but appears to dissociate from Gag prior to budding, although budding rate is decreased and the ABCE1-Gag association is prolonged in cells in which HIV-1 PR is not expressed (33). Despite this information, the role of ABCE1 in HIV-1 assembly is not yet clear.

Large RSV Gag-containing complexes have also been isolated by sedimentation (69).These complexes may correspond to large cytoplasmic aggregates of Gag-GFP observed in the same study, which were thought to contain cellular proteins as well as Gag. The exact sizes and nature of these cellular Gag complexes remain poorly understood.

Cell-based assembly assays are less than ideal for the study of retroviral assembly. While host cell factors required for assembly are best identified and studied in cells, the cell is a complex environment in which any change may affect many apparently independent processes. When studying viral assembly, the number of these potentially confounding processes – transcription, translation, nuclear transit (in some cases), membrane-binding, trafficking, and budding – can be prohibitive. More controlled systems are thus desirable in order to isolate and study retroviral assembly.

## *In Vitro* assembly systems

### Cell-free systems

Two cell-free expression systems have been used extensively for the study of retroviral assembly. The first is an MPMV system based on rabbit reticulocyte extracts. MPMV is a betaretrovirus, so as described above normal cellular MPMV assembly takes place at a cytoplasmic site near the MTOC rather than at the plasma membrane; the cytoplasmic virus-like particles are

then trafficked to the plasma membrane following assembly. The cell-free MPMV assembly system requires the addition of *in vitro* transcribed MPMV mRNA to the extract. The mRNA is translated by ribosomes in the extract, producing, by frameshifting, two products: Gag and Gag-Pro. (Full-length MPMV RNA can undergo two possible frameshifts to produce Gag, Gag-Pro, and Gag-Pro-Pol, but mRNA capable of expressing the full-length Gag-Pro-Pol protein could not be transcribed *in vitro*.) These two proteins assemble into virus-like particles whose density and TEM morphology are comparable to immature MPMV cores. While Gag-Pro was incorporated into particles at wild-type levels, PR activation and maturation did not occur in this system. Other retroviral Gag proteins were not able to assemble in this system (106), suggesting that viruses that pre-assemble in the cytoplasm may have special assembly properties. One such property appears to be the “internal scaffold domain” (ISD), a multimerization domain located in MPMV p12 (between MA and CA). The ISD was required for MPMV assembly in extracts (105) and conferred onto HIV-1 Gag the ability to assemble in this system (104). Multimerization of the ISD is required for assembly to proceed (66). Consistent with the idea that the ISD is an adaptation to pericentriolar assembly, cellular assembly of ISD-mutant Gag can be rescued by re-targeting Gag to the plasma membrane (107).

The second cell-free assembly system, for HIV-1 and other primate lentiviruses, requires a number of cellular co-factors. Full-length HIV-1 Gag mRNA (without Pol coding sequence) is added to wheat germ extract along with myristyl-CoA. Gag is translated and assembles into immature-like particles with a density similar to that of isolated cores (74). Assembly depends on myristylation of Gag, membrane association (membrane

fragments are present in the extract), ATP (74), and a cellular factor called HP68 or ABCE1, a putative RNAse L inhibitor (139). As in cells, assembly appears to proceed through a stepwise series of intermediates identified by sedimentation analysis, although the nature of these intermediates is not clear (74). The same system has been used for assembly of three related primate lentiviruses: human immunodeficiency virus type 2 (HIV-2) and two strains of simian immunodeficiency virus (SIV), SIVmac239, and SIVagm (31). This system would presumably be useful for the assembly of other myristylated Gag proteins.

Cell-free assembly systems have two advantages for the study of retroviral assembly: they have fewer complicating factors than live cells but still support assembly of constructs that are difficult to purify due to low solubility (likely because low expression levels and the presence of cellular chaperones enhance solubility). However, cell extracts still contain a large number of host factors whose effects may be difficult to isolate, and the host factors in the systems used may be significantly different from those found in biologically relevant host cells; for example, wheat germ ABCE has only 71% amino acid identity to human ABCE1 (139). In addition, low expression levels lead to poor particle yields, so cell-free systems do not readily lend themselves to structural analyses. The cell-free systems described thus have some limitations for the study of retroviral assembly.

**Defined-component systems**

Some of the limitations of cell-based and cell-free assembly systems can be overcome by the use of defined systems based on purified components. Such systems allow for the partial separation of the Gag-Gag,

Gag-lipid, and Gag-nucleic acid interactions required for assembly. They can be divided roughly by whether they are based on CA or on full-length Gag (thus producing mature- or immature-like particles); the latter require the addition of nucleic acid and occasionally other constituents, while the former usually do not. HIV-1 CA was the first retroviral Gag protein to be assembled *in vitro*. Under high salt concentrations (0.5-1.0 M) HIV-1 CA assembles into cylindrical or "tubular" particles (36, 45, 72), with occasional conical particles resembling classical HIV-1 cores; some CA mutants appear to favor the conical core morphology (45). Cryo-EM reconstructions of these tubes were among the first evidence for the NTD hexamer/CTD dimer model of the CA lattice (72). Other retroviral CA proteins have proven more refractory to assembly *in vitro*; RSV CA assembles poorly relative to HIV-1 and reports are conflicting regarding the optimal conditions (57, 64). A recently described RSV CA system involving high phosphate concentrations, which produces polyhedral particles resembling authentic cores rather than tubes, is more promising (98). These systems provide a valuable tool for isolating and examining the post-maturation steps of assembly.

*In vitro* assembly systems in which full-length or truncated Gag proteins assemble into immature-like particles have been developed for at least three retroviruses: MPMV, HIV-1, and RSV. MPMV was the first retrovirus to be assembled from purified protein, although the procedure was somewhat complex (65). Full-length Gag protein was expressed in *Escherichia coli*, in which it formed cytoplasmic structures resembling immature particles. The protein was poorly soluble and segregated to inclusion bodies, which also contained partially assembled structures. The inclusion bodies were isolated and solubilized with denaturing concentrations of urea. Renaturation by serial

dialysis produced spherical particles of an appropriate size and density for immature MPMV cores. The nucleic acid requirement was not examined, although it is reasonable to assume that the assembled structures within the inclusion bodies contained bacterial RNA that may have survived denaturation and reassembly. Overall, this system was inconvenient and remains noteworthy only as the first published example of immature retroviral VLPs assembled *in vitro*.

More recently, a second MPMV system was developed to address some of these limitations (120). It was found that CA-NC protein with the N-terminal proline deleted (so as to avoid formation of the β-hairpin and thus presumably retain an "immature" conformation) was highly soluble and readily purified from *E. coli*. This ΔProCANC protein could assemble into aggregates and interlocked spiral structures without nucleic acid or in the presence of short DNA oligonucleotides. Some regular virus-like particles were observed when longer (20-22 nt) oligonucleotides were used. Unlike immature cores, the aggregates migrated to the bottoms of isopycnic sucrose gradients, suggesting higher than normal density. By contrast, abundant regular particles were formed in the presence of long RNA (bacteriophage MS2 or *E. coli* rRNA). The corresponding CA-NC protein (with initial proline intact) formed planar sheets within *E. coli* cells (102) but could not be assembled *in vitro* (120). The effect of the initial proline on particle morphology is intriguing, as it suggests, for MPMV at least, that the absence of the β-hairpin alone may determine immature vs. mature lattice assembly.

The medical significance of HIV-1 has attracted a great deal of interest, and it is thus not surprising that a number of different systems have been developed for *in vitro* assembly of HIV-1 Gag. The oldest of these relies on

HIV-1 MA-CA-NC (the C-terminal p6 domain is deleted as it is otherwise removed by bacterial proteases) expressed in and purified from *E. coli*. Under physiological conditions, this protein assembles into spherical particles in the presence of nucleic acid. With nucleic acid alone, the particles formed are much smaller than authentic immature HIV-1 cores (25-30 nm vs. 100-120 nm) (16); however, when inositol phosphates IP5 or IP6 are added to the reaction, the particles formed are similar in size and morphology to authentic cores (15). This is thus a three-component system, with Gag, nucleic acid, and inositol phosphates required to reconstitute correct immature VLPs.

The inositol phosphate requirement is referable to the MA domain, as the second system, which involves a deletion of amino acids 16-99 of MA, produces normal-sized particles with nucleic acid alone (49). Radial density analysis showed that these particles are similar to authentic immature cores, with the exception of the absent MA domain. This ΔMA construct also has an intriguing pH sensitivity: when assembly is carried out at pH 8.0, the particles formed are all spherical, whereas at pH 6.5 the spheres are mixed with cylindrical particles (50). The nature of this pH-mediated switch is not clear, but a similar sensitivity is involved in monoclonal antibody binding to unassembled Gag, which revealed that epitopes on helices 3 and 6 of the CA NTD are exposed at pH 8.0 but occluded at pH 6.5. This suggests that the pH-mediated morphology switch may be related to a conformational change in Gag (50).

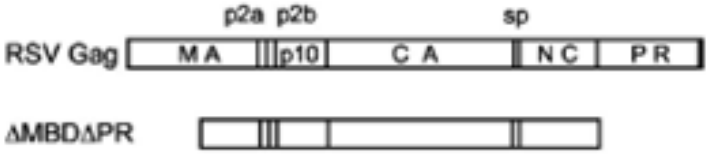
Two other groups have reported *in vitro* assembly of HIV-1 Gag with an intact MA domain and without the addition of inositol phosphates. The first of these systems uses a C-terminally 6His-tagged MA-CA-NC protein. It requires the addition of 5 mM $MgCl_2$ and incubation at 37°C for efficient assembly, and

RNA was reported to be dispensable, although inclusion of rRNA accelerated assembly (85). Assembly efficiency was difficult to interpret and published TEM images suggest that the particle morphology was mildly defective. Assembly appears to proceed through a 60S intermediate (84). The utility of this system remains questionable due to inconsistencies in particle morphology.

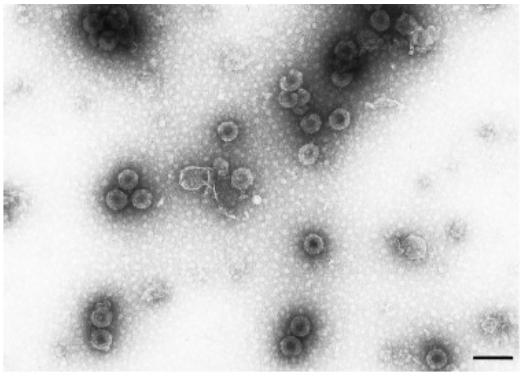
The last of the HIV-1 *in vitro* assembly systems employs an N-terminally His-tagged MA-CA-NC protein (52). Assembly is facilitated by increased temperature (37°C vs. 30°C vs. 23°C); while the addition of RNA reduces the temperature requirement, RNA is not required for assembly. Assembly efficiency was quantified by simple pelleting, which does not distinguish between particles and aggregates, rather than by equilibrium sedimentation which separates particles and aggregates. A pH-sensitive phenotype was reported, with spherical particles observed at pH 8.0 and cylindrical particles at pH 6.0. Particle morphology was subtly unusual in both cases: the spherical particles were packed into a two-dimensional sheet and many lacked the internal staining characteristics of the other preparations, while the cylindrical particles were indistinct. The same group has also reported *in vitro* assembly induced by disulfide-mediated dimerization via C-terminal cysteine residues in place of NC (4), but the particles shown were atypical and this system must currently be regarded as defective at best. The overall lack of consensus on the requirements for HIV-1 Gag assembly *in vitro* probably reflects the number of different researchers and Gag constructs involved.

The third retrovirus for which an *in vitro* immature assembly system has been developed is RSV. The RSV *in vitro* assembly system is based on the truncated Gag construct ΔMBDΔPR, in which the globular membrane-binding

A.




B.



C.

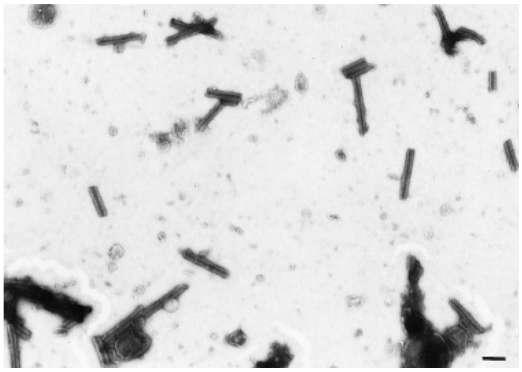


**Figure 1.11**: RSV *in vitro* assembly system. **A.** RSV Gag and ΔMBDΔPR construct used for in vitro assembly. **B.** TEM of spherical particles produced by ΔMBDΔPR. Scale bar represents 100 nm. **C.** TEM of cylindrical particles produced by similar construct lacking p10 domain. Scale bar represents 100 nm. Figure taken from reference (58).

domain of MA and the PR domain of Gag have been removed to improve solubility (17, 133) (**Figure 1.11**). The protein is stored at 500 mM NaCl and pH 8.0. For assembly, the protein is mixed with nucleic acid (10% by weight) and the assembly buffer adjusted to 100 mM NaCl and pH 6.5 by direct dilution or dialysis. Spherical virus-like particles (VLPs) assemble rapidly and efficiently under these conditions, with at least half of the protein assembling into particles. Particles are easily identified and examined by negative staining and transmission electron microscopy. The nucleic acid requirement is absolute, but a wide variety of nucleic acids (RNA, ssDNA, and dsDNA), as well as other polyanions such as heparin sulfate, support assembly (133). The smallest ssDNA oligonucleotide that allows assembly appears to be equivalent to two NC binding sites, suggesting that dimerization may be sufficient to initiate assembly in this system (79). Radial density analysis confirms that the particles produced are, allowing for protein truncation, identical to immature cores produced from cells (133). The requirement for the C-terminal 25 amino acids of p10 for spherical assembly was first identified in this system (58) (**Figure 1.11**). The RSV *in vitro* assembly system remains the most robust and possibly the most deeply studied of the *in vitro* immature assembly systems.

While defined component assembly systems typically rely on non-physiological protein concentrations and lack host factors that might facilitate assembly at lower concentrations, they do have advantages relative to the cell-based and cell-free systems previously described. Because these systems are based on purified Gag proteins, the effects of other components (e.g. nucleic acid, inositol phosphates, or, hypothetically, host cell proteins) can be readily examined. The absence of cell debris means that particle morphology can be examined by simple negative staining and TEM, rather than requiring

laborious thin-sectioning; this makes *in vitro* assembly a useful screening procedure for Gag mutations that may cause assembly defects. The isolation of assembly from other processes involved in virus-like particle production is useful when trying to determine when a particular defect in particle production occurs. Finally, the large amounts of protein produced in *E. coli*, and the purity of the particles or assembly intermediates produced, make these systems suitable for in depth structural study.

## ONGOING QUESTIONS IN RSV ASSEMBLY

### Relevance of extended NTD structure to immature assembly

The mechanism by which the C-terminal 25 amino acids of RSV p10 promote spherical assembly was not known at the inception of the work presented in this thesis. A possible clue to the function of this sequence came with the publication of a crystal structure of the N-terminal domain of RSV CA with a 25 amino acid upstream extension (91). This extended NTD protein crystallized as a dimer with an extensive interface between p10 and the NTD (**Figure 1.7**). The correspondence between this structure and the HIV-1 conformational switch is intriguing. In the latter, switching from the sphere-promoting to the tube-promoting pH occludes epitopes on helix 3 and helix 6 of the N-terminal domain of CA (50). The analogous helices in the RSV NTD are involved in the extended NTD dimer interface. However, modeling this structure into a virus particle requires significant bending of CA at the flexible linker (91), whereas radial density studies on spherical VLPs assembled *in vitro* and on immature cores imply that Gag is extended with a linear array of domains (133). In addition, a random mutagenesis of the critical region of p10

produced normal (spherical) and cylindrical particles in *E. coli* lysates, as well as proteins that failed to assemble, but there was no discernable pattern to the assembly results, let alone any correspondence to the crystal structure (58). Therefore, the biological significance of this structure was in doubt, and the mechanism of the p10 morphology switch remained unknown at the inception of this thesis research.

The p10 domain of RSV Gag is also implicated in nuclear export of Gag in cells. In cultured cells, an MA-GFP fusion protein, unlike the Gag-GFP fusion, localizes to the nucleus, while extension through p10 restores cytoplasmic localization. This implies that there is a nuclear export signal (NES) within p10. The NES was subsequently localized to amino acids 24-57, overlapping with the 25 residues required for spherical assembly, amino acids 33-57. This NES was found to be dependent on the cellular nuclear export factor CRM1 based on nuclear accumulation of Gag following treatment with leptomycin B, an inhibitor of the CRM1-dependent nuclear export pathway. CRM1-dependent nuclear export sequences are typically leucine-rich, and there is such a motif (**L**XX**W**XX**V**XXX**L)** within the last 25 amino acids of p10. Some mutant RSV virus constructs that package only a single copy of the viral genome (instead of the dimeric genome required for infectivity) fail to accumulate in the nucleus when treated with LMB, and it is hypothesized that nuclear transit is required for packaging of a dimeric genome (111). Unfortunately, the dual requirement for these residues for nuclear export and assembly has complicated analysis of both functions (112). As part of my thesis work, I performed a systematic investigation of the assembly function of the last 25 amino acids of RSV p10, using the *in vitro* assembly system to

bypass the NES function, in order to elucidate both the assembly function of this critical region.

**Initiation of assembly by dimerization**

Dimerization is a critical step in retroviral assembly in cells, as demonstrated by the production of particles when the NC domain of HIV-1 is replaced by a dimeric “leucine zipper” (1, 137); the same result is seen for RSV in a baculovirus/Sf9 expression system, with normal particle morphology confirmed by electron microscopy (56). Dimerization is also critical *in vitro*: while diverse nucleic acids support assembly, there is a minimum size that corresponds to two NC-binding sites (79). Both dimerizing and trimerizing leucine zippers have been used to drive HIV-1 assembly *in vitro* (Alan Rein, personal communication). Replacement of HIV-1 NC with a single cysteine residue that is oxidized to form disulfide-mediated dimers also leads to assembly *in vitro* (4).

No artificial dimerization system has proven successful for RSV assembly *in vitro*. This may be due to a functional element in RSV Gag, recently identified by this laboratory, which spans the C-terminal tail of CA, the SP domain, and the first four amino acids of NC. We have seen that disruption of this sequence by insertion or substitution produces tubular Gag particles in the baculovirus expression system in insect cells(60); parallel constructs fail to assemble *in vitro* (59). All of the prior attempts to introduce artificial dimerization domains have involved fusing the domain in question directly to the C-terminus of SP, removing the last four amino acids of the CA-SP-NC element. While such a fusion formed spherical particles in the baculovirus expression system, it is possible that the degree of overexpression inherent to

that system masks some subtle defect in assembly. An *in vitro* assembly system driven by artificial dimerization would help to clarify the roles of dimerization and of the CA-SP-NC element in RSV assembly.

## THESIS OUTLINE

In this dissertation, I describe a series of experiments undertaken to explore the roles of the p10 domain and of dimerization in RSV assembly. As related in Chapter 3, I began by performing a systematic mutagenesis of the last 25 amino acids of p10 in which I identified six amino acids that abrogated *in vitro* assembly when individually mutated to alanine. Of these six amino acids, five made side chain contacts with the CA domain in the extended NTD structure, while 14 of the 15 non-alanine amino acids that were dispensable for assembly made no such contact. In addition, mutation of two amino acids in helix 3 of CA that made hydrogen bonds with the p10 domain in the extended NTD structure also abrogated assembly. To determine whether the p10-CA interface actually existed in virus-like particles, I introduced a pair of cysteine residues in place of E51 of p10 and T20 of CA. These exogenous cysteines were predicted by the extended NTD structure to form a disulfide bond across the p10-CA interface. I then developed an *in vitro* oxidation protocol that allowed me to demonstrate specific p10-CA cross-linking in VLPs assembled *in vitro* and in cells. The product of cross-linking across the interface was not a dimer, as predicted by the extended NTD structure, but a high molecular weight complex. I isolated the high MW complex by sedimentation and used partial reduction to demonstrate that it was in fact a hexamer. Collaboration with a molecular modeling group produced models of

the mature and immature RSV NTD hexamers which agree with recently obtained electron cryotomography data for immature retroviral particles.

As described in Chapter 4, I next attempted to develop a system for assembling RSV Gag by artificial dimerization, employing a variety of Gag-leucine zipper fusion proteins as well as an FKBP-based inducible dimerization system developed by Ariad Pharmaceuticals. I first took a Gag-leucine zipper fusion that had produced morphologically normal particles in a baculovirus overexpression system and expressed it in avian cells. This construct was able to bud from avian cells but did not do so reproducibly. I then attempted to perform *in vitro* assembly of Gag-leucine zipper fusion proteins using heterodimeric zipper pairs and recently obtained knowledge of an upstream assembly element to improve on previous failed attempts. Both homo- and heterodimeric leucine zipper fusions failed to dimerize (by sedimentation analysis) or assemble (by TEM) *in vitro*. Attempts to assemble Gag-FKBP fusion proteins in cells were also unsuccessful: neither budding from cells nor sedimentation *in vitro* showed any response to the dimerizing reagent. I was not able to determine whether this was due to a failure of the protein to dimerize or to dimerization in the absence of the reagent. Future directions for this line of research are described in Chapter Five.

# CHAPTER TWO

# MATERIALS AND METHODS

## DNA constructs for p10 mutagenesis and cross-linking studies

The pET3xc.ΔMBDΔPR plasmid has been described previously (17). pET3xc.ΔMBDΔPR (ΔBlpI) was generated by partially digesting ΔMBDΔPR with BlpI, gel-purifying single-cut plasmid, blunting the ends with Taq polymerase, and ligating to destroy the site; clones were screened by XhoI/BlpI double digestion to identify one with a unique BlpI site in CA. This step destroyed a BlpI site in the vector backbone and left a unique BlpI site in the NTD of CA. pET3xc. ΔMBD(-6C) is an unpublished plasmid (constructed by May Ma (77)) in which the six cysteines in NC were replaced by alanine by two-step PCR. ΔMBD(-11C) was made by replacing the XhoI/FseI fragment of ΔMBD(-6C) with a PCR product in which the five cysteines upstream of p10 were mutated to serine by two rounds of two-step PCR. Mutations involving P38, G39, or P40 of p10 were made by replacing the XhoI/BlpI fragment of ΔMBDΔPR(ΔBlpI) with two-step PCR products containing the desired mutation; the remainder of the alanine and GSGSG substitutions, and E51C, T20C, and E51C/T20C, were made by replacing the FseI/BlpI fragments of ΔMBDΔPR(ΔBlpI) with two-step PCR products containing the desired mutation(s). E51C(-11C), T20C(-11C), and E51C/T20C(-11C) were made by replacing the FseI/SacII fragment of ΔMBD(-11C) with the analogous fragment from the analogous ΔMBDΔPR(ΔBlpI) constructs. TWCC(-11C) was made by replacing the FseI/SacII fragment from ΔMBD(-11C) with a two-step PCR product amplified from E51C/T20C(-11C). P38CC(-11C) was made by replacing the XhoI/FseI fragment of T20C(-11C) with a two-step PCR product

that used ΔMBD(-11C) as a template. E51Cf was made by replacing the FseI/BlpI fragment of ΔMBDΔPR(□BlpI) with a two-step PCR product using p10(58-62)f as a template. E51C-f-T20C was made by replacing the FseI/BlpI fragment of ΔMBDΔPR(ΔBlpI) with a 2-step PCR product using E51Cf as a template to make E51C-f-T20C. Finally, E51C-f-T20C(-11C) was made by replacing the FseI/SacII fragment of ΔMBD(-11C) with that from E51C-f-T20C.

The protease pET11c.ePR(C113S) was made by replacing the AvrII fragment of pET11c.his-ePR (110) with two annealed oligonucleotides.

pQCXIP.SMGagΔPR(-11C) was made by replacing the XhoI-PacI fragment of the unpublished plasmid pQCXIP.SMGag (constructed by Amanda Dalton) with a PCR product amplified from ΔMBD(-11C). The other cell expression constructs were then made by replacing the XhoI/BlpI fragment of SMGagΔPR(-11C) with that from the analogous *E. coli* expression vector.

All PCR-derived fragments and cloning sites were confirmed by sequencing.

**DNA constructs for dimerization studies**

The pCMV.MA-CA-LZ and pCMV.MA-CA constructs were made by replacing the BglII-NotI fragment of pEGFP.N2 (Clontech) with a double-stranded oligonucleotide containing a KpnI site. The SacI/KpnI fragment of the intermediate plasmid was replaced with the SacI/KpnI fragment from pFB.MA-CA-Zip or pFB.MA-CA (56). The pCMV.MA-CA-FKBP plasmid was made by replacing the SacII/KpnI fragment of pCMV.Gag-D37S (69) with a PCR product amplified from pC4-$F_v$1E (Ariad Pharmaceuticals). The forward primer was extended at the 5' end to contain the Gag sequence from the SacII site through nt 1464 of Gag so that the SP/NC border of Gag was fused to amino

acid 5 of $F_v1$. The remainder of the $F_v1$ domain through the stop codon was included in the PCR product.

The *in vitro* dimerization constructs were made by inserting an HpaI site into pET3xc. ΔMBDΔPR(ΔBlpI) at nt 1227 of the Gag sequence to make pET3xc. ΔMBD-NC4-HpaI. This required two conservative nucleotide substitutions and resulted in HpaI cutting at the NC+4 position without changing any amino acids. The HpaI/KpnI fragment of pET3xc. ΔMBD-NC4-HpaI was replaced with primer extension products to generate pET3xc. ΔMBD-zip, pEt3xc. ΔMBD-Kzip, pET3xc. ΔMBD-Zip3, and pET3xc. ΔMBD-Zip12. The HpaI/KpnI fragment was replaced with a PCR product amplified from pC4-$F_v1E$ to generate pET3xc. ΔMBD-NC4-fFKBP, in which amino acid 4 of NC was fused to a GSGSG linker followed by amino acid 6 of $F_v1E$ (amino acid 5 of $F_v1E$ is a glycine). A 3C protease cleavage site (EVLFQ/GP) followed by a 6-His tag was included at the C-terminus of the $F_v1E$ sequence.

**Protein purification**

ΔMBDΔPR and derivatives containing an NC domain were expressed and purified as previously described (133) with the following changes: after ammonium sulfate precipitation, the protein was further purified by cation exchange chromatography using an Amersham AKTA *prime* FPLC system. The ammonium sulfate pellet was dissolved in 10 mL 20 mM Tris pH 8.0, bound to an Amersham SP FPLC column in 20 mM Tris pH 8.0, 100 mM NaCl and eluted with 20mM Tris pH 8.0 and 500 mM NaCl in a single step. DEAE and phosphocellulose purification and dialysis steps were omitted. Protein concentration was assayed using Advanced Protein Assay reagent (Cytoskeleton, Inc.) and adjusted to 5 mg/mL by centrifugal filtration (5000

MWCO) or dilution with 20 mM Tris pH 8.0 and 500 mL NaCl. DTT (10 mM) was added to proteins which were not to be used for cross-linking studies.

ΔMBDΔPR proteins lacking an NC domain were purified as follows. ΔMBD-Kzip had a strongly positively-charged C-terminal domain and was purified as though it contained NC. ΔMBD-bzip, ΔMBD-Zip3, and ΔMBD-Zip12 were purified similarly to ΔMBDΔPR through the ammonium sulfate precipitation step. The ammonium sulfate pellet was resuspended in 5 mL 20 mM Tris pH 8.0, 100 mM NaCl and allowed to flow through both DEAE and phosphocellulose columns (~2 mL) followed by a 5 mL wash with 20 mM Tris pH 8.0, 100 mM NaCl. The flowthrough and washes were pooled and dialyzed 4 hours to overnight against 20 mM Tris pH 8.0, 500 mM NaCl, and 10 mM DTT. Proteins were concentrated to 5 mg/mL by centrifugal filtration (5000 MWCO). Protein concentration assays and storage conditions were the same as for ΔMBDΔPR.

ΔMBD-NC4-fFKBP was purified by nickel affinity chromatography. Protein was expressed in *E. coli*, the cells were lysed, and the lysates were cleared as for ΔMBDΔPR. The ultracentrifugation supernatant was incubated with 2.5 mL 50% Ni-NTA agarose (Qiagen) rocking for 1 hour at 4°C. The beads were washed with 2 X 10 mL 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole. The protein was eluted with 4 X 1 mL 20 mM Tris pH 8.0, 500 mM NaCl, and 250 mM imidazole. The majority of the protein was contained in the second elution fractions. This fraction was dialyzed overnight against 20 mM Tris pH 8.0, 500 mM NaCl, and 10 mM DTT. Protein concentration assays and storage conditions were the same as for ΔMBDΔPR.

RSV ePR and ePR (C113S) were purified as previously published (110) with minor modifications. Following sonication, the proteins were insoluble and

were purified from inclusion bodies. Inclusion bodies were collected by centrifugation at 10,000 x g for 15 minutes, resuspended and re-pelleted twice with 20mM Tris pH 7.5, 2M urea, 2% Triton X-100, 10 mM DTT, 1 mM PMSF and a third time with the same buffer without urea or Triton X-100, and pelleted again. The pellet was dissolved in 10 mL 20 mM Tris pH 7.5, 7 M urea, and 10% glycerol and stirred for 30 minutes at room temperature. Insoluble material was pelleted (10 min at 10K rpm in SA-600) and discarded. The supernatant was dialyzed overnight against 20 mM Tris pH 7.5, 20 mM NaCl, 1 M urea, and 10% glycerol. Precipitated protein was removed by centrifugation (10 min at 10K rpm in SA-600). The supernatant was incubated with 2.5 mL 50%  Ni-NTA slurry rocking for 1 hour at 4°C. The beads were washed with 2 X 10 mL 20 mM Tris pH 8.0, 20 mM NaCl, 1M urea, 20 mM imidazole, 10 mM DTT. The protein was eluted with 4 X 1 mL elution buffer (20 mM Tris pH 8.0, 20 mM NaCl, 1 M urea, 250 mM imidazole, 10 mM DTT). The majority of the protein was found in the second elution fraction. This protein fraction was dialyzed overnight against 20 mM Tris pH 6.0, 100 mM NaCl, 1% glycerol, and 0.4 M urea. Precipitated protein was removed as previously described. The supernatant was incubated at 37°C for 5 h to overnight to allow self-cleavage of the 6-His tag. The reaction was dialyzed against the same buffer to remove the cleaved His tag and insoluble material was removed as previously described. The protein concentration was assayed by UV spectroscopy and the protein was stored at -80° C.

### *In vitro* assembly and electron microscopy

Proteins containing NC were assembled by dilution to 1 mg/mL in 50 mM MES pH 6.0 or 6.5 with or without 50 μM DTT, addition of 10% wt/wt

ssDNA (GT50, an oligonucleotide with 25 GT repeats (133)), gentle mixing, and incubation at room temperature for 15 minutes. Proteins without NC were diluted to 1 mg/mL in 50 mM MES pH 6.5 without GT50, with the exception of ΔMBD-Kzip, which was assembled with 10% GT50. ΔMBD-NC4-fFKBP was assembled by diluting to 1 mg/mL with 50 mM MES pH 6.5 and adding a stoichiometric amount (10 μM) of AP20187.

Assembly reactions were spotted onto carbon-Formvar-coated grids and stained with 2% uranyl acetate (pH 5.0). The particles were visualized on a Morgagni 268 transmission electron microscope. Proteins scored as assembly-competent yielded many particles visible in a single field. Proteins scored as assembly-incompetent yielded no regular particles.

**Centrifugation**

Equilibrium centrifugation was performed on 10-60% (wt/wt) sucrose gradients in assembly buffer (50 mM MES pH 6.5 and 100 mM NaCl). Gradients were spun in an SW60 rotor at 50,000 rpm for 4 h at 4°C. Fractions containing VLPs were collected from the gradient and their densities measured with a refractometer. The VLP fractions were diluted 1:5 with assembly buffer and the VLPs were pelleted by spinning in a microfuge at 16,000 x g for 30 min at 4°C.

Rate-zonal centrifugation was performed on 10-30% wt/wt sucrose gradients. The gradients were made in 50 mM MES pH 6.5, 500 mM NaCl, 2 M urea for the cross-linking experiments and in 50 mM Tris pH 8.0, 100 mM NaCl for the dimerization assays. Gradients were spun in an SW60 rotor at 50,000 rpm for 24h at 4°C.

**Oxidation and PR digestion**

Particles were resuspended in assembly buffer with or without 10 μM DTT. Resuspended particles were oxidized with copper phenanthroline (60 μM $CuSO_4$, 267 μM o-phenanthroline), mixed by vortexing for approximately five seconds, and immediately quenched with 20 mM iodoacetamide and 3.7 mM Neocuproine (Sigma). *In vitro* PR digestion was carried out as previously described (110). In brief, reactions were adjusted to 100 mM MES pH 6.5, 700 mM NaCl, 0.7 mM EDTA, and 5.3% glycerol. Prior to digestion, DTT to 15 mM was added to reduced samples and fresh iodoacetamide was added to oxidized samples. A cysteine-lacking mutant of RSV PR (PR C113S) purified from *E. coli* was added at 2 μg PR/10 μg ΔMBDΔPR. (Wild-type PR is inactive in the presence of iodoacetamide.) Reactions were incubated at 37°C for 16h. Fresh iodoacetamide was added to oxidized samples immediately prior to SDS-PAGE.

**Cell culture, transfection, and transduction**

DF1 (chicken embryonic fibroblast) cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (or 5% fetal bovine serum and 5% NuSerum (BD Biosciences)), 1% heat-inactivated chick serum, standard vitamins, L-glutamine, penicillin and streptomycin.

For MA-CA-LZ and MA-CA-FKBP experiments, cells were maintained in 60 mm dishes and transfected with 4 μL FuGENE transfection reagent (Roche) and 1 μg DNA per dish.

For cross-linking experiments, stable cell lines expressing Gag proteins were created by transduction with virus derived from Phoenix cells transfected with pQCXIP.SMGagΔPR(-11C) and related constructs. Phoenix cells were

transfected using FuGENE 6 (Roche). 24 hours post-transfection, the medium from the Phoenix cells was collected, filtered (sterile 0.45 μM syringe filters), mixed 1:1 with fresh medium and applied to DF1 cells. If DF1 cells appeared vacuolated or otherwise unhealthy, the medium was replaced with 100% fresh DMEM after 12 hours. This infection procedure was repeated daily until the Phoenix cells began lifting off of the dish or the DF1 cells became confluent (2-3 days). Transduced cells were selected with 2 μg/mL puromycin until widespread apoptosis was evident (~3 days).

### $^{35}$S labeling and immunoprecipitation

24-48 hours post-transfection, cells were labeled with 100 mCi $^{35}$S methionine/cysteine in 500 μL methionine/cysteine-free DMEM per 60 mm dish for 15 minutes. Separate dishes were labeled for each construct at each planned timepoint. After the 15 minute labeling period, cells were rinsed X 3 with PBS and 2 mL DMEM was added to each dish for the chase period.

At the designated timepoints, medium was removed and virus-like particles were collected as described below. Cells were rinsed with PBS and incubated with 0.5 mL RIPA buffer (20 mM Tris pH 7.5, 150 mM NaCl, 1% deoxycholic acid (sodium salt), 1% Triton X-100, 1 mM PMSF) for ten minutes on ice. Dishes were scraped with rubber policeman or cell lifter, lysates were transferred to 1.5 mL Eppendorf tubes, and dishes were washed with an additional 0.5 mL RIPA buffer which was combined with the lysates. Lysates were spun at 10,000 X g (11,000 rpm in microcentrifuge) for five minutes to pellet the nuclei. The supernatants were transferred to new tubes and mixed with 1 μL each rabbit anti-RSV MAp10 and rabbit anti-RSV CA serum and incubated rotating 1-16 hours at 4°C. 35 μL Protein A-Sepharose beads (1:1

slurry) was added to each lysates and incubated rotating 1-2 h at 4°C. The beads were washed with 3 X 1 mL RIPA buffer with 15 minute incubations between each wash. Finally, the beads were boiled for ten minutes with 40 µL SDS-PAGE sample buffer.

**Virus-like particle collection from cell culture medium**

Medium was spun at 1500 X g (~4000 rpm in a microfuge or ~3000 rpm in a clinical centrifuge) for five minutes to pre-clear. Supernatants were loaded into polyallomer TL 100.4 tubes. For some $^{35}S$ labeling experiments, purified unlabeled ALV virus was added as a carrier. Approximately 0.5 mL 15% sucrose (wt/wt) in virion suspension buffer (VSB) (20mM Tris pH 7.5, 100 mM NaCl, 1mM EDTA) was added to the bottom of each tube as a cushion. Samples were spun at 70K rpm for 20 minutes in the TL100.4 rotor. Medium was aspirated off and the sides of the tube were rinsed with PBS or VSB before the sucrose cushion was aspirated off as well.

**Autoradiography**

For 35S labeling experiments, pelleted VLPs were resuspended in SDS-PAGE sample buffer and 100% of both immunoprecipitation and pelleted VLP samples were loaded for SDS-PAGE. Following electrophoresis, the gels were dried and exposed to a $^{35}S$ phosphor screen for 24-48 hours (or longer as needed). The screen was read using a STORM 860 phosphorimager (Molecular Dynamics) and gel files were analyzed using ImageQuant software.

**Immunoblotting**

For cross-linking experiments, confluent transduced cells were split 1:8 and cells and medium were harvested 3 days later. Cells were trypsinized, rinsed with PBS, and resuspended directly in SDS-PAGE sample buffer. 1% of each cell lysate was loaded for SDS-PAGE. VLPs were collected from the medium as described above. For cross-linking experiments, pellets were resuspended in VSB plus 20 mM iodoacetamide and permeabilized with 0.01% Triton X-100 in the same buffer. PR digestion was carried out as previously described. 10% of each sample was loaded for SDS-PAGE. Following SDS-PAGE, gels were transferred to nitrocellulose membrane (30 minutes- hour at 100 V, 20% methanol in transfer buffer). For cross-linking experiments, the membrane was incubated shaking in “block” (10% non-fat dried milk in TBST (20 mM Tris pH 8.0, 137 mM NaCl, 0.1% Tween-20)) for 16 hours at 4°C. The membrane was washed 2 X five minutes with TBST and incubated with the primary antibody (polyclonal rabbit anti-RSV CA serum 1:1000 in block) for one hour at room temperature. The membrane was washed for ten minutes with block and 2 X ten minutes with TBST and incubated with the secondary antibody (monoclonal anti-rabbit IgG conjugated to alkaline phosphatase (Sigma), 1:30,000 in block) for one hour at room temperature. The membrane was washed 3 X twenty minutes in TBST. ECF reagent (Amersham) was applied to the membrane, and the fluorescence was immediately read using the Storm 860 phosphorimager (Molecular Dynamics) on the blue fluorescence setting. The gel files were analyzed using ImageQuant software.

For FKBP budding experiments, cells and medium were harvested 24 hours post-transfection. VLPs were collected as described above. Cells and

VLP pellets were resuspended directly in SDS-PAGE sample buffer and 1% of lysates and 10% of VLP pellets were loaded for SDS-PAGE. Immunoblotting was performed as described above with the following modifications. The first incubation of the membrane in block was performed for one hour at room temperature. The primary antibody was diluted in TBST rather than in block and hybridization was performed for 16 hours at 4°C. The secondary antibody was also diluted in TBST rather than in block.

**Molecular modeling**

Molecular modeling of the RSV NTD hexamers was performed by Paul Murray (Columbia University). The structures for the MLV CA NTD hexamer (PDB code 1U7K, (88)) and the RSV extended NTD dimer (PDB code 1P7N, (91)) were used to model the mature RSV CA NTD hexamer. The protein secondary structure prediction programs, PSIPRED (82), 3D-PSSM (61), and Scratch (97) predict a β-hairpin in residues 1-15 of RSV CA NTD, as seen in the mature RSV CA NTD structure (PDB code 1EM9, (64)). Consequently, the modeling program Nest (96) was used to construct a composite model of RSV CA NTD in which residues 1-21 were modeled after 1U7K and 22-144 were modeled after 1P7N. The resulting mature monomer model was then structurally super-imposed, using the Combinatorial Extension (CE) server (115), on each sub-unit (A-F) of 1U7K, the mature MLV CA NTD hexamer, thus yielding a mature model of the RSV CA NTD hexamer.

The same structures (1U7K & 1P7N) were used to construct an immature model of the RSV CA NTD hexamer. As opposed to the mature CA, immature CA is not cleaved between Met0 and Pro1; i.e., the upstream p10 helix is joined to and free to interact with CA. In order to model the immature

hexamer, the a-B inter-molecular pair of p10-CA NTD from 1P7N was excised. The p10 half, “a”, was cut after CA Thr6; therefore, p10 comprises Pro-24 to Thr6 of “a”. The CA NTD half, “B”, was cut before Leu14; therefore, CA NTD comprises Leu14 to Ala144 of “B”. The “a-B” pair was then structurally super-imposed, using CE, on each sub-unit of 1U7K. With the “p10” region added in, there was significant structural overlap between the sub-units of the immature hexamer. In order to relieve this overlap, each sub-unit was radially moved away from the center of the hexamer until there was no structural overlap. The program Loopy (96) was used to model in the missing segment between residues Thr6 and Leu14. This segment was modeled such that each p10 region is joined to the next CA NTD of the hexamer, moving clock-wise.

# CHAPTER 3

# RELEVANCE OF THE EXTENDED NTD STRUCTURE TO IMMATURE ASSEMBLY[1]

## INTRODUCTION

Rous Sarcoma Virus is a well-studied avian virus that is frequently employed as a model retrovirus. Among the advantages of RSV as a model is a robust *in vitro* system for immature assembly. A truncated RSV Gag termed ΔMBDΔPR (**Figure 1.11**) can be purified and assembled *in vitro* into spherical virus-like particles (VLPs) that are similar in shape, size, and radial density profile to immature cores produced from cells (17, 133). Deletion of p10, the Gag domain immediately upstream of CA, produces cylindrical particles (17); as purified HIV-1 CA assembles into cylinders *in vitro* (36, 45), cylindrical morphology is considered indicative of a mature (or mature-like) CA lattice. Internal and N-terminal deletions of p10 give the same result, so the morphology change cannot be ascribed to NTD β-hairpin formation alone. The spherical assembly determinant was further localized to the C-terminal 25 amino acids of p10, both *in vitro* (58) and in a baculovirus expression system (56). The same sequence also contains a CRM-1-dependent nuclear export signal (NES) required for proper intracellular trafficking of Gag and for viral genome packaging (13, 111-113); the significance of this dual function of p10 is unknown.

[1] Expanded from **Phillips, J.M., P.S. Murray, D.Murray and V.M. Vogt**. 2008. A molecular switch required for retrovirus assembly participates in the hexagonal immature lattice. Embo J **27**:1411-1420.

When the RSV NTD was extended N-terminally to include the 25 amino acids immediately upstream (the last 25 amino acids of p10), it crystallized as an anti-parallel dimer with an extensive p10-CA interface (91) (PDB ID: 1P7N) (**Figure 1.7**). The extended NTD monomer is similar in structure to the mature NTD previously characterized by NMR and crystallography (19, 64, 91) but lacks the N-terminal β-hairpin of CA, which is destabilized when proline 1 of CA is not free to form the salt bridge with aspartate 52 in helix 3 (**Figure 3.1**). The dimer interface includes residues in a short α-helix in the p10 segment. Because of the anti-parallel arrangement of the NTDs, the extended NTD dimer presumably cannot be a subunit of an NTD hexamer and therefore is probably not found in the immature virus particle. In addition, assembly results from a random mutagenesis of the critical region of p10 did not obviously correspond to the structure (58). That said, the requirement of the same 25 amino acids of p10 for both spherical virus particle morphology and the extended NTD dimer interface suggests that the p10-CA contacts in the dimer interface may have some relevance to immature assembly.

## RESULTS

### Systematic mutagenesis suggests extended NTD structure is relevant to immature assembly

I sought to determine whether the extended NTD structure was relevant to immature assembly using the RSV *in vitro* assembly system. I chose the *in vitro* system for this study for two reasons. First, the *in vitro* system enabled me to study the assembly function(s) of the C-terminal 25 amino acids of p10 without regard to the overlapping NES discussed above and in Chapter 1.

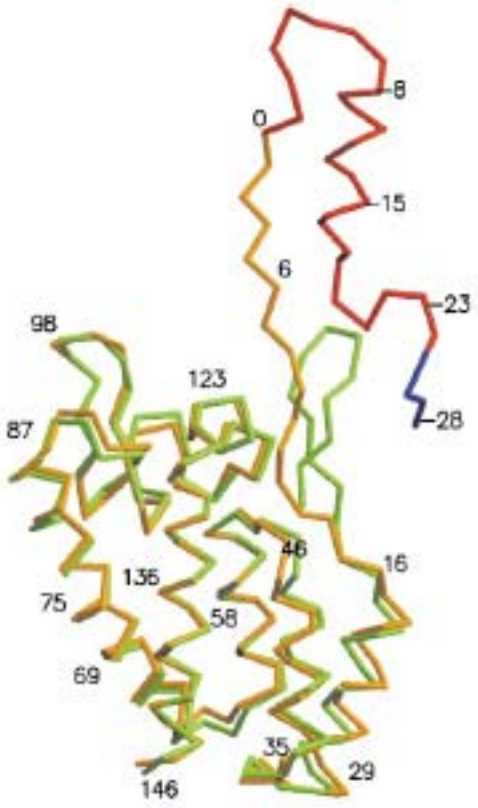



**Figure 3.1**: Extended NTD is largely similar to the mature NTD. The mature NTD backbone (64) is in green. The extended NTD is in gold (CA residues) and red (p10 residues). The backbones overlap well except in the β-hairpin region, which is unfolded in the extended NTD structure. Figure taken from reference (91).

Second, examination of *in vitro* assembled particle morphology by TEM requires only negative staining, a process that is faster and simpler than fixation and thin-sectioning of particle-producing cells. Because the results of a previous random mutagenesis of the region were inconclusive (58), I first performed a systematic alanine mutagenesis in the context of the *in vitro* assembly construct ΔMBDΔPR (**Figure 1.11**). The critical 25 amino acid region was divided into five segments of five amino acids and each segment was mutated in turn to either AAAAA or GSGSG. The resulting constructs were expressed in *Escherichia coli*, purified, assembled *in vitro* as described in Chapter Two, and screened for assembly by negative staining and TEM. I performed two independent expressions and purifications of each protein. All proteins showed a consistent phenotype, yielding either abundant regular spherical particles or no regular particles. Segments of five amino acids that could be mutated to AAAAA or to GSGSG without affecting assembly were scored as dispensable for assembly. Segments of five amino acids for which mutation to both AAAAA and GSGSG prevented assembly were subdivided into smaller alanine substitutions until all individual amino acids required for assembly were identified. The four endogenous alanine residues were not individually mutated.

The results of this mutagenesis (**Table 3.1**) showed that the last 10 amino acids of the critical region allowed extensive substitution, while the first 15 amino acids of the critical region contained a mix of required and dispensable amino acids. While the last 10 amino acids were dispensable for assembly, the mutagenesis results showed that they were not all functionally equivalent in this assembly assay: the most C-terminal segment (residues 58-62; PVVAM) tolerated mutation to GSGSG but not to AAAAA, while the

**Table 3.1**: Constructs used to test p10-NTD interface. “A” indicates that the indicated residues were mutated to alanine, “f” that the indicated residues were mutated to GSGSG.

| **p10 mutants** | **Assembly phenotype** |
|---|---|
| (38-42)A | - |
| (43-47)A | - |
| (48-52)A | - |
| (53-57)A | + |
| (58-62)A | - |
| (38-42)f | - |
| (43-47)f | - |
| (48-52)f | - |
| (53-57)f | + |
| (58-62)f | + |
| (38-39)A | + |
| P40A | + |
| L42A | - |
| T43A | - |
| D44A | - |
| W45A | - |
| (47-48)A | - |
| R47A | + |
| V48A | - |
| (49-50)A | + |
| (51-52)A | - |
| E51A | + |
| L52A | - |
| **CA mutants** | |
| D52A | - |
| N55A | - |
| **Combination mutants** | |
| (T43, W45)A/(53-57)A | - |
| (T43, W45)A/(58-62)f | - |
| (T43,W45,V48,L52)A | - |
| (T43,W45)A/N55A | - |

preceding segment (residues 53-57; ASTGP) tolerated both mutations. These results suggested that while none of the side chains in the last 10 amino acids of p10 is required for assembly, there are some sequence constraints on the final five amino acids. By contrast, some of the first fifteen amino acids of p10 appear to make specific side chain contacts that are required for assembly.

There are two amino acids in the NTD that make predicted side chain hydrogen bonds across the dimer interface in the extended NTD structure, CA D52 and N55, both of which are located in helix 3 of CA. Mutation of either of these to alanine also abrogated assembly (**Table 3.1**). In addition, I attempted to obtain cylindrical particles by making various combinations of p10 and NTD mutations; unfortunately, none of these constructs produced any regular particles (**Table 3.1**).

Comparison of the mutagenesis results with the published amino acid contacts in the extended NTD interface (91) revealed a striking pattern. Of the six amino acids in p10 whose side chains contacted the NTD in the structure, five were required for assembly and only one (E51 in the p10 helix) was dispensable (**Figure 3.2**). Similarly, of the fifteen non-alanine amino acids whose side chains were not involved in the interface, fourteen were dispensable and only one (D44 upstream of the p10 helix) was required. These results strongly support the hypothesis that the p10-CA interface in the extended NTD structure is involved in immature assembly.

### The p10 domain of Gag can be cross-linked to CA via disulfide bonds

The ability of the E51A mutant to assemble despite altering a side chain contact with the NTD offered an opportunity to use disulfide cross-linking to

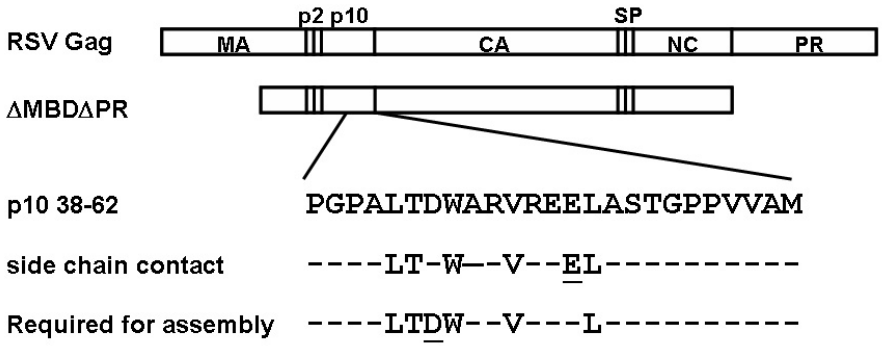



**Figure 3.2**: *In vitro* assembly results correlate with side chain contacts in extended NTD structure. C-terminal 25 amino acids of RSV p10 domain with mutagenesis results. “Side chain contact” indicates those amino acid residues whose side chains contact the CA domain in the extended NTD structure. “Required for assembly” indicates those amino acid residues which, when changed to alanine, abrogate assembly *in vitro*.

test for the presence of the extended NTD p10-CA interface in assembled virus particles. The Swiss-PDBViewer program (51) predicted a disulfide bond when E51 in the p10 helix and the residue it contacts in CA, T20 in helix 1, were mutated to cysteines (**Figure 3.3**) in the context of the extended NTD structure. Therefore, the p10-CA interface depicted in the 25NTD structure could be detected in virus particles by mutating these residues to cysteine and assaying for disulfide cross-linking between p10 and CA.

Initial experiments conducted with air oxidation showed promise, but cross-linking was inefficient, possibly due to residual dithiothreitol (DTT) from the protein purification procedure. To improve cross-linking, I developed a chemical oxidation protocol using copper phenanthroline as the oxidizing agent, *p*-chloromercuribenzoic acid (PCMB) or iodoacetamide as a "quenching" reagent to decrease non-specific cross-linking of oxidized cysteines, and neocuproine as a scavenger to inactivate the excess copper. (PCMB caused a noticeable band shift in some assays and was eventually replaced with iodoacetamide.) This protocol dramatically improved cross-linking efficiency but led to extensive non-specific cross-linking of the 12 endogenous cysteines in the ΔMBDΔPR protein (data not shown). In order to reduce this interference, I mutated the cysteines in MA, p2, and NC to alanine or serine (**Figure 3.4a**). One endogenous cysteine, C192 in the C-terminal domain of CA, was known from previous studies to disrupt assembly when mutated to serine (77) and was therefore left in place. The resulting construct, ΔMBD(-11C) [hereafter called -11C] assembled *in vitro* into normal spherical particles (**Figure 3.4b**). All subsequent cysteine mutations were made in the context of -11C. Mutation of p10 E51 [E51C], CA T20 [T20C], or both [CC] to cysteines allowed spherical assembly *in vitro* (**Figure 3.4b**), although the

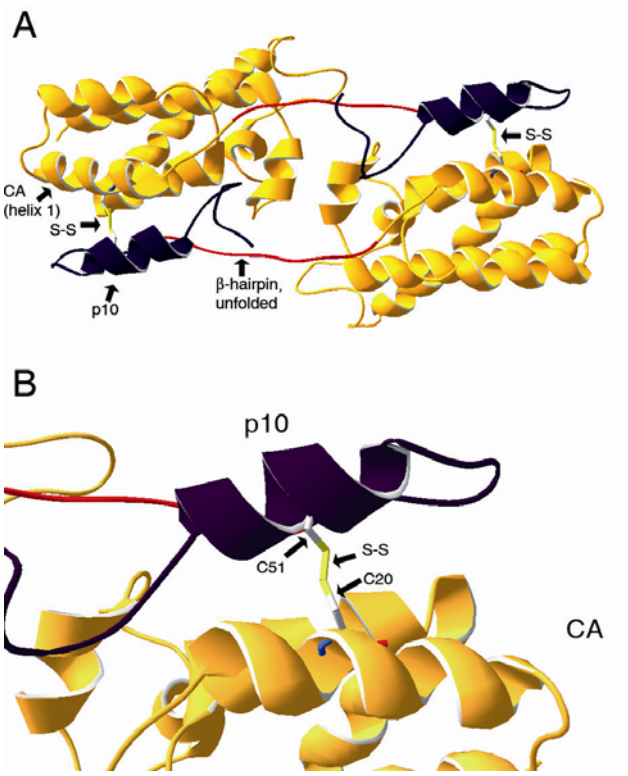



**Figure 3.3**: Extended NTD structure predicts disulfide cross-linking when E51 of p10 and T20 of CA are mutated to cysteines**(A)** High-resolution structure of the extended NTD dimer (1P7N) showing disulfide bonds predicted to form when p10 E51 and CA T20 are mutated to cysteines. The disulfide bonds (yellow) link CA (orange) and p10 (purple) domains across the dimer interface. Those residues that form the β-hairpin in mature CA are shown in red. **(B)** Close-up of boxed area in **A** showing the predicted E51C-T20C disulfide bond.

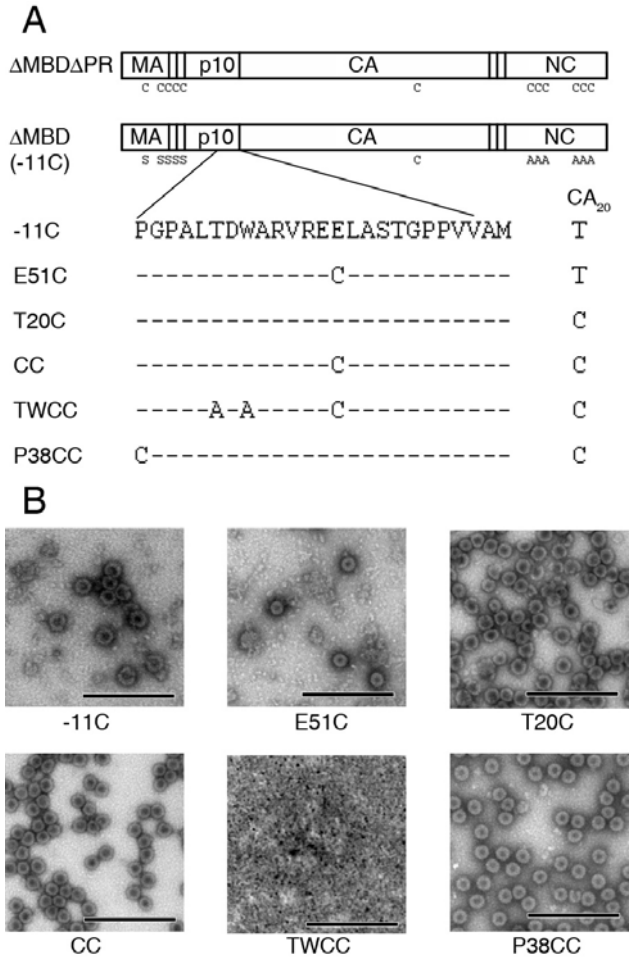



**Figure 3.4**: Cysteine mutant constructs used for cross-linking. **(A)** Schematic representation of ΔMBDΔPR showing locations of the endogenous cysteine residues and the cysteine mutations used in this study. **(B)** Transmission electron microscopy images of ΔMBDΔPR cysteine mutants assembled *in vitro* and negative stained. -11C, top left; E51C, top middle; T20C, top right; CC, bottom left; TWCC, bottom middle; P38CC, bottom right. Scale bars represent 500 nm.

E51C single mutant assembled less efficiently. Both -11C and E51C were found to assemble more efficiently at pH 6.0 than at pH 6.5, and some assemblies of these proteins were performed at the lower pH. As a negative control for assembly, the p10 T43A and W45A mutations, which abrogate assembly (**Table 3.1**), were combined with the CC construct to produce TWCC (**Figure 3.4a**); as predicted, TWCC did not assemble (**Figure 3.4b**), confirming that the cysteine mutations could not rescue a p10-CA interface defect.

In initial experiments, I performed cross-linking on unassembled protein (protein under assembly conditions with no nucleic acid added) and on total assembly reactions (DNA oligonucleotide GT50 added at 10% wt/wt). Samples were incubated with and without RSV PR to separate the Gag domains and thus identify the site(s) of cross-linking. Digested samples were then analyzed by non-reducing SDS-PAGE. Potential outcomes of these cross-linking experiments are diagrammed for reference in **Figure 3.5**.

Upon analysis, -11C protein produced bands consistent with monomeric full-length (50 kDa) and CA (27 kDa) proteins under all conditions (**Figure 3.6a,** lanes 1-3), suggesting that C192 in CA does not form disulfide bonds under these experimental conditions. E51C protein produced three bands in the absence of PR (**Figure 3.6a**, lane 4): a monomer, an upshifted monomer (due to PCMB modification of p10; data not shown), and a dimer. As the E51C CA was monomeric (**Figure 3.6a**, lanes 5-6), the dimer was presumably due to p10-p10 cross-linking (see **Figure 3.5**, example 1). Full-length T20C protein and T20C CA were also predominantly monomeric (**Figure 3.6a**, lanes 7-9), suggesting that C20 alone does not form either homotypic or heterotypic (with C192) disulfide bonds. Full-length CC protein

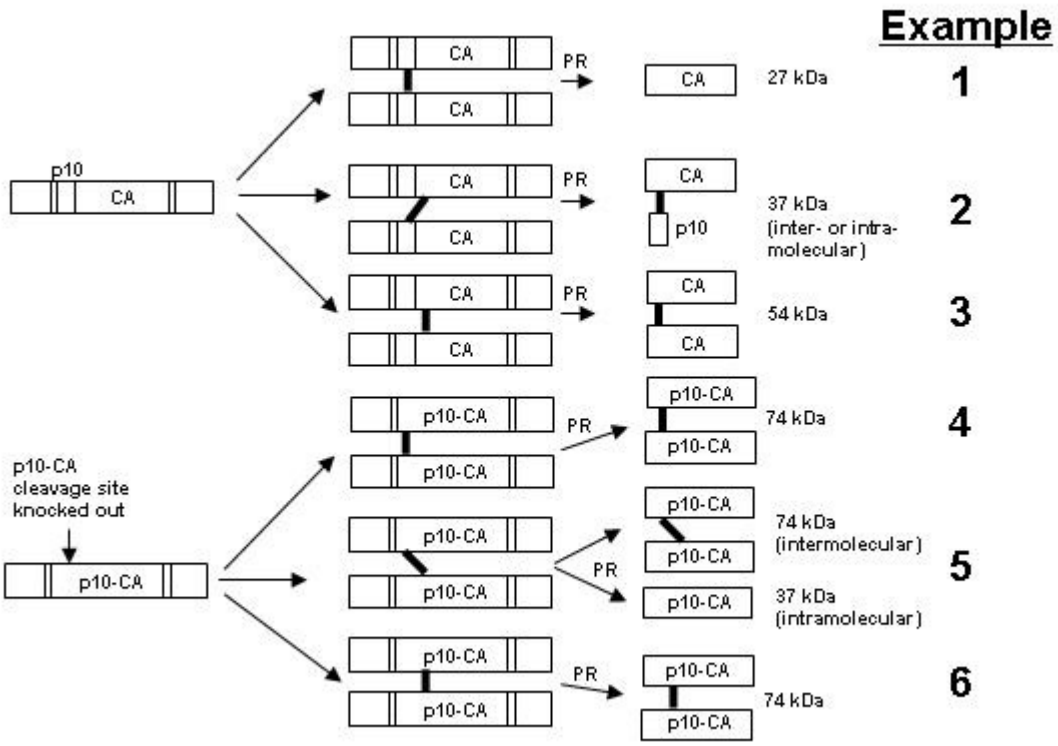



**Figure 3.5**: Possible outcomes of disulfide cross-linking of cysteines in p10 and CA when analyzed by non-reducing SDS-PAGE after PR cleavage. Heavy black lines represent disulfide bonds.

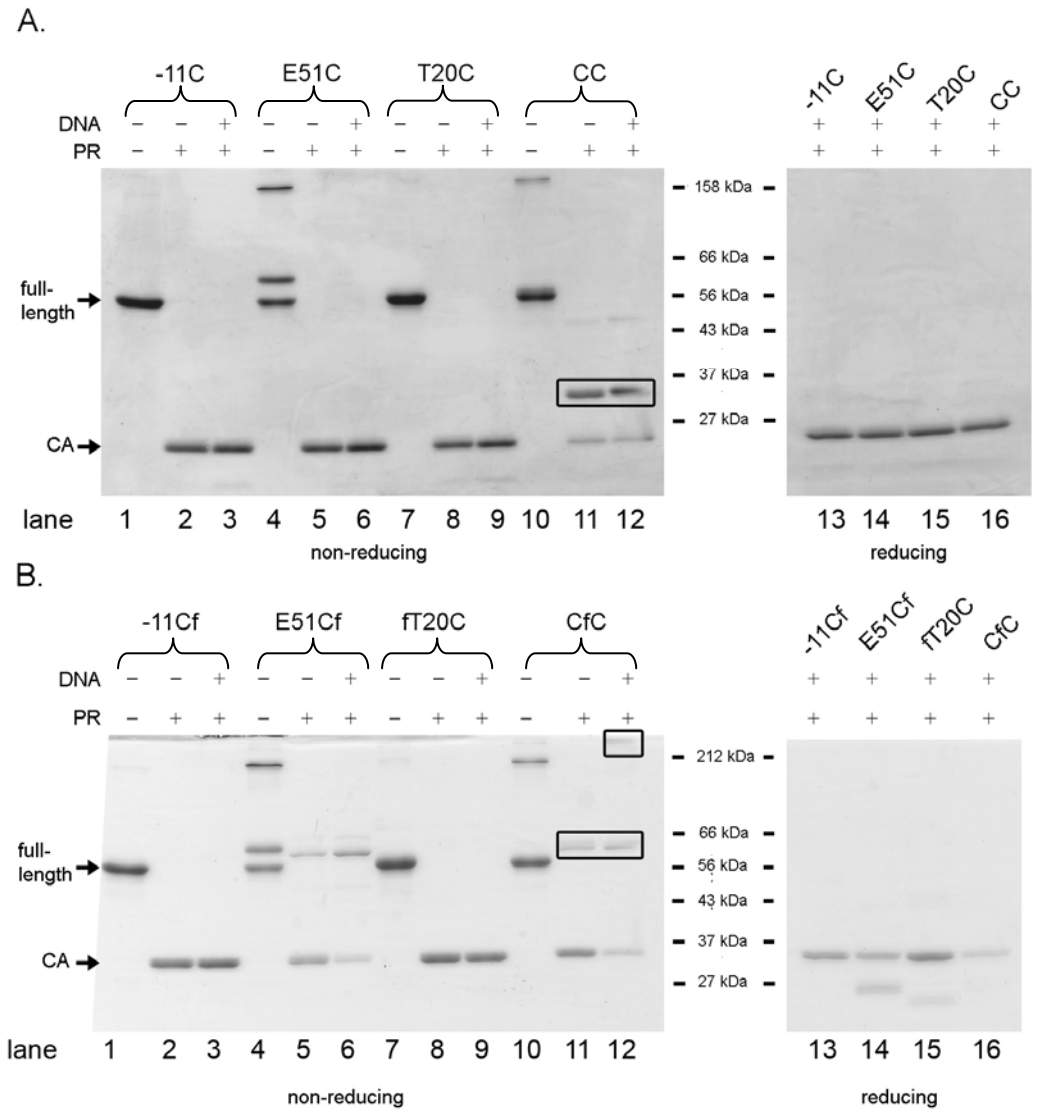



**Figure 3.6**: p10-CA cross-linking is seen upon introduction of cysteines predicted to form disulfide bonds across the p10-CA interface. A. Constructs with wild-type PR cleavage sites. Addition of nucleic acid to reaction and PR to digests is noted above each lane. Samples on the far right are reduced. Note novel 37 kDa band in lanes 11-12. B. Constructs with p10-CA cleavage site knocked out. Note p10-CA dimers in lanes 5-6 and 11-12 and high MW band in lane 12.

produced both monomer and dimer bands (**Figure 3.6a**, lane 10), while PR digestion yielded three distinct CA bands (**Figure 3.6a**, lanes 11-12): a monomer, a 37 kDa band (boxed) as predicted for p10-CA cross-linking (**Figure 3.5**, example 2), and a very faint, approximately 50 kDa band. The 50 kDa band might represent a CA dimer (**Figure 3.5**, example 3), although this seems unlikely as no comparable band was produced by digested T20C protein. More likely, the 50 kDa band results from incomplete PR digestion of the cross-linked product. (In this and subsequent experiments, cross-linked CC protein was difficult to digest completely under non-reducing conditions.) The pattern for PR digested CC protein was the same whether or not nucleic acid was added, suggesting that the cross-linking seen did not depend on the presence of assembled particles.

To confirm that higher species were due to disulfide bond formation, I also digested assembled samples under reducing conditions and analyzed them by standard reducing SDS-PAGE. All samples reduced to monomeric CA products (**Figure 3.6a**, lanes 13-16), indicating that higher species were due to disulfide cross-linking. Immunoblotting using constructs with an internally HA-tagged p10 confirmed that the 37 kDa band observed for digested CC protein contains both p10 and CA (**Figure 3.7**). These results confirm disulfide bonding between the exogenous cysteine residues C51 of p10 and C20 of CA.

To distinguish between intra- and inter-molecular cross-linking, I compared PR-digested samples from constructs with wild-type cleavage sites to samples from parallel constructs in which the p10-CA cleavage site was abolished by combining the cysteine mutations with the p10(58-62)f mutation (p10(58-62)f) allows assembly (**Table 3.1**) but disrupts PR cleavage). The

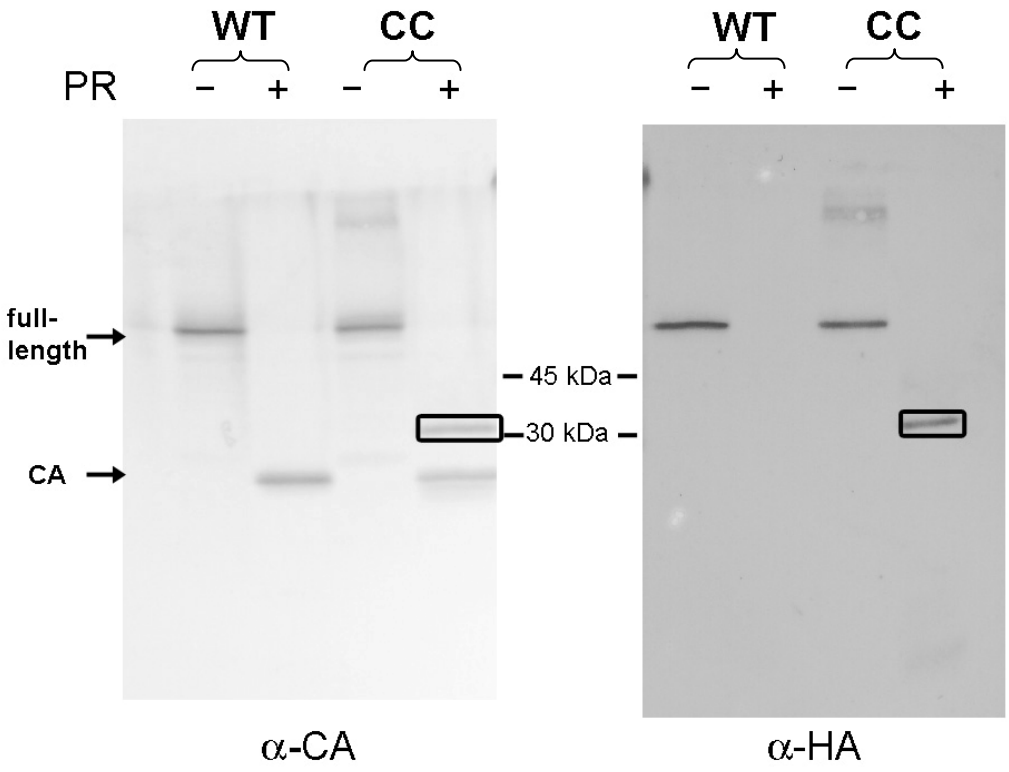



**Figure 3.7**: Novel 37 kDa band contains p10 and CA. -11C and CC constructs were made with an internal HA tag in p10. Unassembled proteins were oxidized and quenched as described and analyzed by immunoblotting with antibodies to CA and HA. Note that 37 kDa band is recognized by both antibodies.

combination yielded constructs -11Cf, E51Cf, fT20C, and CfC. When analyzed by non-reducing SDS-PAGE as before, intermolecular p10-CA cross-linking of the CfC construct should yield a 37 kDa band when the cleavage site is intact (**Figure 3.5**, example 2) but a 74 kDa band when cleavage is disrupted (**Figure 3.5**, example 5), whereas intramolecular cross-linking should yield a 37 kDa band in either case (**Figure 3.5**, examples 2 and 6).

The -11Cf, E51Cf, and fT20C constructs produced monomeric full-length and 37kDa p10-CA bands as predicted (**Figure 3.6b,** lanes 1-9). E51Cf also produced an upshifted full-length band (**Figure 3.6b**, lane 4) due to PCMB modification (data not shown) and dimeric p10-CA bands (**Figure 3.6b**, lanes 5-6), due presumably to p10-p10 cross-linking (**Figure 3.5**, example 4). In addition to the bands seen for E51C, CfC produced a third, very high molecular weight species, but only when nucleic acid was added (**Figure 3.6b**, lane 12). All high molecular weight species, including the very large species produced by CfC, disappeared under reducing conditions, confirming that the high molecular weight species were produced by disulfide cross-linking.

Based on these results, I initially hypothesized that the dimeric intermolecular p10-CA interface seen in the extended NTD structure did not exist in virus-like particles but represented an intramolecular (monomeric) p10-CA interface found in both unassembled Gag and in the virus particle. I attributed the p10-p10 dimers to transient interactions between unassembled proteins and the very high molecular weight species produced by assembled, digested CfC to non-VLP aggregates formed during assembly.

**Cross-linked Gag in particles is multimeric**

To test my initial hypothesis that the p10-CA interface is intramolecular in virus-like particles, I subjected cross-linked, non-digested CfC assembly reactions to equilibrium sedimentation analysis on 10-60% sucrose gradients. Under these conditions, particles are expected to migrate to approximately fraction 12 (45% sucrose, density = 1.22), while unassembled proteins remain at the top of the gradient and aggregates migrate to the bottom of the gradient. Surprisingly, the fractions that contained VLPs contained predominantly high-molecular weight cross-linked species (**Figure 3.8**, boxed bands and TEM inset). This suggested that the p10-CA interface in virus-like particles is intermolecular rather than intramolecular and that the species trapped by cross-linking p10 and CA in VLPs is much larger than a dimer.

**p10-CA cross-linking is intramolecular and non-specific in solution**

Disulfide cross-linking of Gag proteins can occur in situations other than stable contacts between folded proteins in virus-like particles, including transient interactions between proteins in solution, protein-protein interactions in misassembled aggregates, and interactions between denatured proteins in SDS-PAGE preparations. The probability of such non-specific cross-linking can be decreased by quenching any remaining cysteines with an excess of a sulfhydryl-reactive reagent prior to denaturation with SDS and by purifying VLPs prior to cross-linking, but these tactics are not 100% effective. To distinguish between interface-specific and non-specific p10-CA cross-linking, I made an additional construct by mutating p10 P38 at the beginning of the critical region (**Figure 3.2**) to cysteine and combining it with T20C to make P38CC (**Figure 3.4a**). P38 was chosen because it allows assembly when

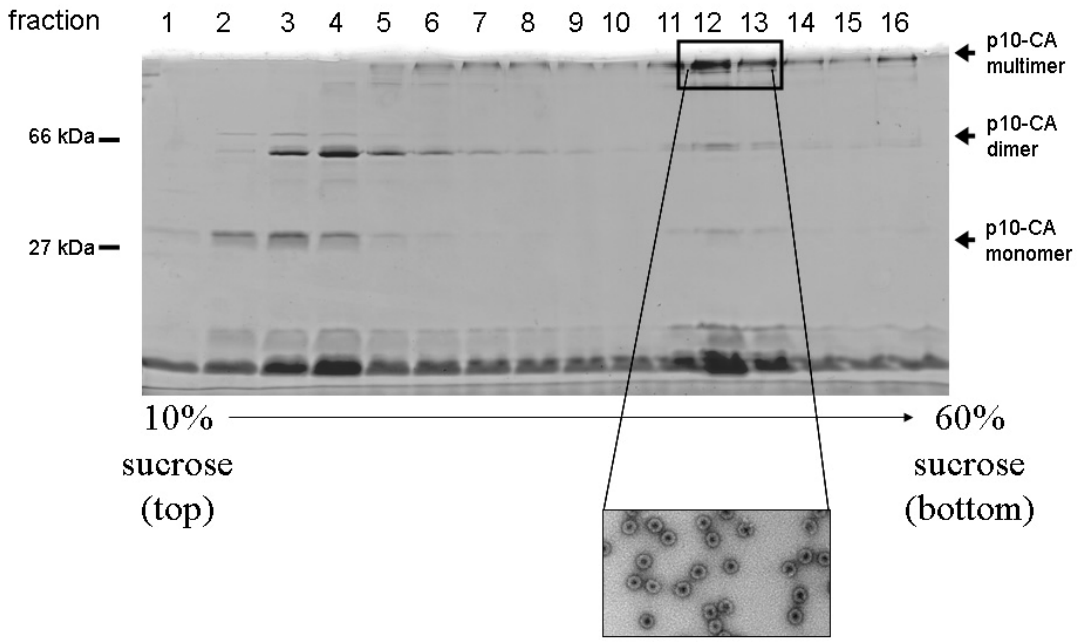



**Figure 3.8**: Gag in virus-like particles is multimeric. Total assembly reaction using CfC construct was analyzed by equilibrium sedimentation on 10-60% sucrose gradient. VLPs were visible in the gradient and were found in fractions 12-13 (TEM inset). Fractions were digested with PR and subjected to SDS-PAGE. Note that cross-linked protein in lanes 12-13 is predominantly multimeric.

mutated to alanine (**Table 3.1**) but is over 15Å away from T20 of the other subunit in the extended NTD structure, making disulfide formation between cysteines at these sites impossible in the context of the p10-CA interface depicted in the crystal structure. P38CC should thus allow non-specific p10-CA cross-linking (during transient collisions or when proteins are denatured) but should not support cross-linking if the p10-CA interface depicted in the extended NTD structure is present. The P38CC protein assembled as expected (**Figure 3.2b**) and was used in the following experiments as a specificity control for the presence of the p10-CA interface.

I first examined the specificity of p10-CA cross-linking in unassembled proteins by oxidizing under assembly conditions without nucleic acid. As before, I incubated samples with and without RSV PR to localize cross-linking and analyzed the products by non-reducing and reducing SDS-PAGE. The results are shown in **Figure 3.9**. In the absence of PR or reducing agent, all constructs were predominantly monomeric with the exception of E51C (**Figure 3.9**, top left, lane 2), which was largely dimeric. The dimers are due to p10-p10 disulfide cross-linking, as the CA band is monomeric upon digestion (**Figure 3.9**, bottom left, lane 2) and the full-length dimer disappears upon reduction (**Figure 3.9**, top right, lane 2). All three constructs with exogenous cysteines in both p10 and CA show a 37 kDa band upon PR digestion (**Figure 3.9**, bottom left, lanes 4-6). As the full-length proteins are monomeric (**Figure 3.9**, top left, lanes 4-6) and the digested bands reduce to 27 kDa (**Figure 3.9**, bottom right, lanes 4-6), the 37 kDa band is most likely due to intramolecular p10-CA disulfide cross-linking (**Figure 3.5**, example 6). As the CC construct appears identical to both controls (TWCC, which is not predicted to form the p10-CA interface, and P38CC, which is not predicted to cross-link across the

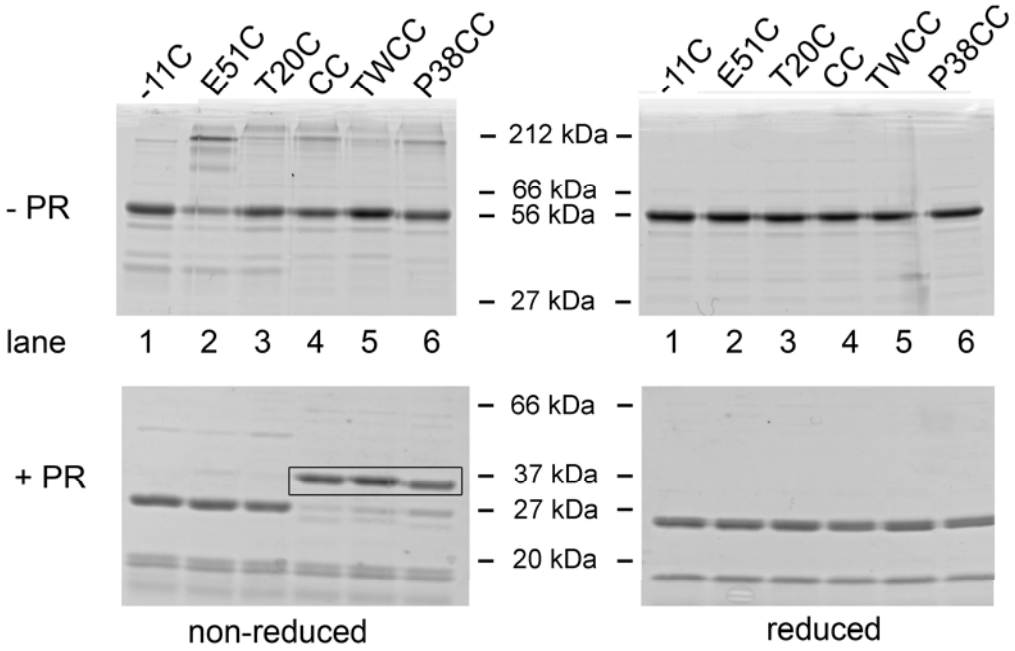



**Figure 3.9**: Unassembled proteins cross-link non-specifically. Samples were oxidized, quenched, and treated as described in the text, before SDS-PAGE. Lane 1, -11C; lane 2, E51C; lane 3, T20C; lane 4, CC; lane 5, P38CC. In the bottom panel the samples were treated with PR before SDS-PAGE. Samples on left were not reduced. Samples on the right were reduced before PR digestion and SDS-PAGE. The 37 kDa band appears in lanes 4, 5, and 6 on the lower left.

interface), I concluded that unassembled protein forms intramolecular p10-CA disulfide bonds in a non-specific manner. This experiment also validated the specificity control P38CC: the P38CC protein demonstrates non-specific p10-CA cross-linking in the absence of the p10-CA interface depicted in the extended NTD structure.

### p10-CA cross-linking is intermolecular and specific in *in vitro* assembled virus-like particles

To test for the presence of the p10-CA interface in assembled particles, I purified virus-like particles from assembly-competent constructs (therefore excluding TWCC) by equilibrium sedimentation, oxidized the particles (with or without stoichiometric amounts of DTT), and quenched as described in Chapter 2. I incubated the oxidized particle samples with and without recombinant RSV PR to separate the Gag domains and analyzed the products by non-reducing and reducing SDS-PAGE. The results are shown in **Figure 3.10**. The -11C protein was predominantly monomeric for both full-length protein (**Figure 3.10**, top left, lane 1) and CA (**Figure 3.10**, bottom left, lane 1), and neither band shifted upon reduction (**Figure 3.10**, top and bottom right, lane 1), indicating that C192 in the CTD does not form homotypic disulfide bonds under these assay conditions. Some of the full-length E51C protein (**Figure 3.10**, top left, lane 2) was dimeric under non-reducing conditions but reduced to a monomer (**Figure 3.10**, top right, lane 2), while the CA was monomeric even when oxidized (**Figure 3.10**, bottom left, lane 2), suggesting that the full-length dimers were due to p10-p10 disulfide bonds (**Figure 3.5**, example 1). Notably, there was no evidence that C51 in p10 was able to cross-link to C192 in CA. Full-length T20C protein (**Figure 3.10**, top left, lane

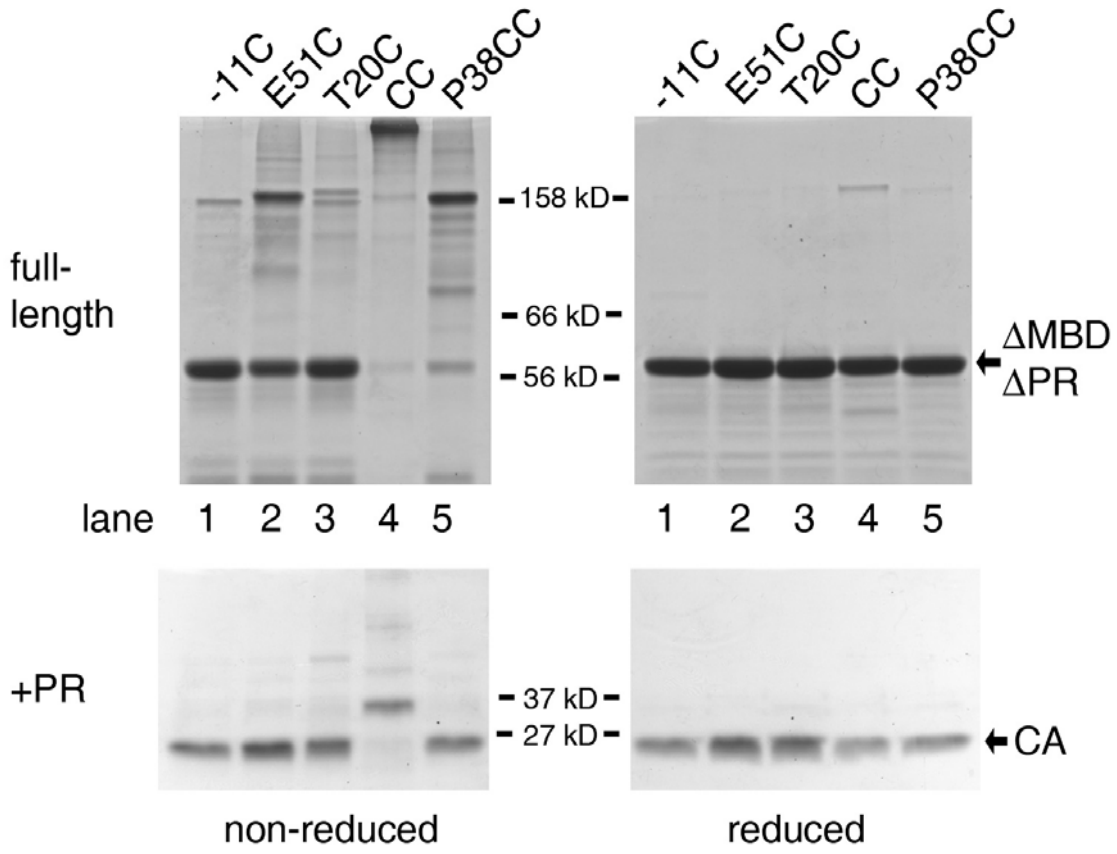



**Figure 3.10**: Exogenous cysteines in p10 and CA form specific, intermolecular disulfide bonds in virus-like particles. Particles were oxidized, quenched, and treated as described in the text, before SDS-PAGE. Lane 1, -11C; lane 2, E51C; lane 3, T20C; lane 4, CC; lane 5, P38CC. In the bottom panel the samples were treated with PR before SDS-PAGE. Samples on left were not reduced. Samples on the right were reduced before PR digestion and SDS-PAGE. The novel high molecular weight band appears in lane 4 on top left, and the novel shifted CA band appears in lane 4 on bottom left.

3) also formed disulfide-mediated dimers, although to a lesser extent. These must be CA-CA dimers as there are no cysteine residues in any other Gag domain, and upon PR digestion a faint corresponding upper band is present at approximately 50 kDa (**Figure 3.10**, bottom left, lane 3) as predicted for a CA dimer (**Figure 3.5**, example 3). It is not possible to distinguish whether these dimers are mediated by C20-C20 or C20-C192 disulfide bonds, and both species may well be present; in any case, the majority of the T20C protein is monomeric. The results for these single mutants suggest that neither of the exogenous cysteines, C51 in p10 or C20 in CA, produces extensive cross-linking when introduced alone.

In contrast to the preceding control proteins, full-length CC protein (**Figure 3.10**, top left, lane 4) formed a distinct, very high MW band with almost no monomer or dimer present, while the CA band (**Figure 3.10**, bottom left, lane 4) was shifted upward by approximately 10 kDa with no 27 kDa CA remaining. Both full-length and CA bands reduced to the normal monomer position (**Figure 3.10**, top and bottom right, lane 4), confirming that the band shifts observed were due to disulfide bonds. For the specificity control P38CC, the full length protein was largely dimeric (**Figure 3.10**, top left, lane 5), but the CA was predominantly monomeric with a very faint upper band similar to that seen for T20C (**Figure 3.10**, bottom left, lane 5). This suggests that most of the full-length dimers were due to p10-p10 disulfide bonds mediated by C38 (**Figure 3.5**, example 1), with a few CA-CA disulfides as for T20C (Figure 3.5, example 3); again, there was no evidence of p10-CA cross-linking, indicating that C38 in p10 was unable to form disulfide bonds with either C20 or C192 in CA. This was in marked contrast to the results seen for oxidized unassembled protein, in which either C38 or C51 in p10 was able to cross-link to C20 in CA

(**Figure 3.9**), and suggests that the measures taken to purify VLPs prior to oxidation and to quench remaining cysteines prior to denaturation were largely successful in eliminating non-specific p10-CA cross-linking. Together, the cross-linking results indicate that the p10-CA interface depicted in the extended NTD crystal structure is present in virus-like particles but that the cross-linked species is much larger than the predicted dimer.

### p10-CA cross-linking is intermolecular and specific in cell-produced virus-like particles

To ensure that these results did not represent an artifact of *in vitro* assembly or of chemical oxidation, I made parallel versions of the constructs described above in an MLV-based retroviral vector expressing RSV Gag with the PR domain removed. Because the RSV nuclear export signal is contained within the region of p10 under study and includes the W45 residue, it seemed likely that the TWCC construct, and possibly others as well, would be trapped in the nucleus and would therefore be unable to assemble and bud. To avoid this problem, all cell expression constructs included a modified MA domain called Super M which contains two mutations in the membrane-binding domain of MA, E25K and E70K. The net +4 charge is reported to increase the affinity of Gag for the plasma membrane and thereby to bypass nuclear transit (14); combining Super M MA with the constructs described in **Figure 3.4a** was predicted to ensure that all Gag proteins reached the plasma membrane.

I created DF1 (chicken embryonic fibroblast) cell lines stably expressing these constructs by MLV transduction. Virus-like particles produced by the cell lines were collected and quenched as described in Chapter Two, permeabilized with mild detergent treatment to disrupt the viral envelopes,

incubated with and without PR, and analyzed by reducing and non-reducing SDS-PAGE and immunoblotting. All proteins were expressed (data not shown) and all constructs except for TWCC were able to assemble and bud from cells (**Figure 3.11**). Gag was monomeric or dimeric and CA was the correct size in particles produced from all constructs except for CC(-11C), which produced a high molecular weight Gag complex and an upward-shifted CA, exactly as seen *in vitro* (**Figure 3.11**, top left, lane 4). I concluded that the interaction between p10 and CA occurs in membrane-enclosed VLPs produced from cells as well as in particles assembled with purified protein *in vitro*, and that efficient disulfide bond formation can occur by air oxidation.

**The p10 and CA domains of Gag become cross-linked as a hexamer**

To identify the high MW band produced by undigested CC, the high molecular weight product was isolated and partly reduced to form a ladder. Due to the amount of protein required, the parallel construct with the mutated p10-CA cleavage site, CfC(-11C), was used. As described above, this protein combines the CC construct with the p10(58-62)GSGSG mutation, which allows assembly but destroys the p10-CA cleavage site. In the absence of PR, CfC behaves similarly to CC, but it can consistently be purified with a much higher yield. I assembled, collected and oxidized VLPs as previously described and then resuspended the particles in 500 mM NaCl plus 2M urea to disrupt weak non-covalent intermolecular interactions. The resulting sample was subjected to rate-zonal centrifugation in a sucrose gradient to separate protein complexes by size, and the fractions were analyzed by non-reducing SDS-PAGE (**Figure 3.12**). Two distinct protein complexes were observed: a major complex that migrated significantly more slowly than 212 kDa and a

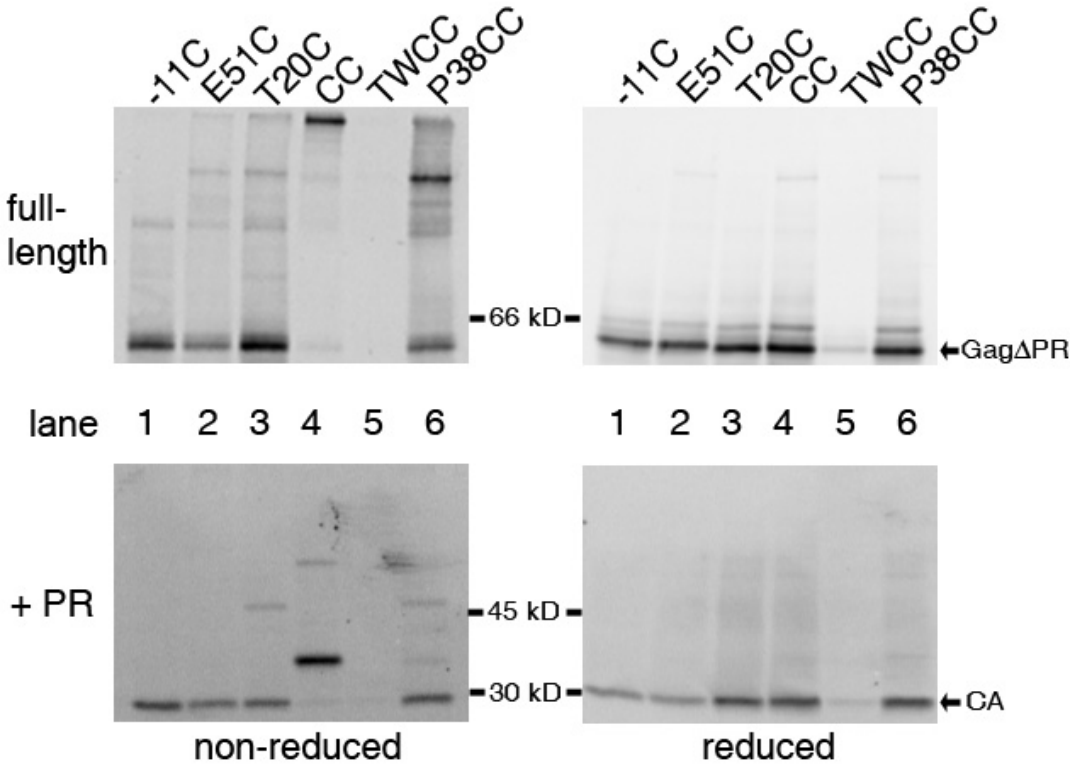



**Figure 3.11**: Specific intermolecular p10-CA cross-linking occurs on air-oxidation of virus-like particles produced from cultured avian cells. Particles were quenched and treated as described in the text before SDS-PAGE and immunoblotting with rabbit anti-CA serum. Lane 1, -11C; lane 2, E51C; lane 3, T20C; lane 4, CC; lane 5, TWCC; lane 6, P38CC. Samples in the bottom panels were treated with PR before SDS-PAGE. Samples on left were not reduced. Samples on the right were reduced before PR digestion and SDS-PAGE. The novel high molecular weight band appears in lane 4 on top left, and the novel shifted CA band appears in lane 4 on bottom left.

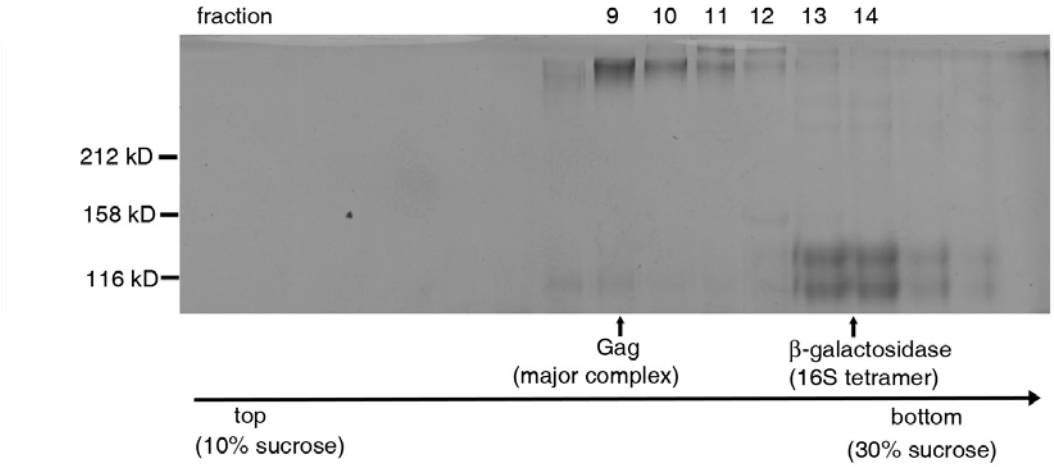



**Figure 3.12**: High MW Gag complex contains two species. Analysis by rate-zonal centrifugation. Cross-linked particles were dissolved in 2M urea and 0.5M NaCl, mixed with *E. coli* β-galactosidase as an internal size standard, and then centrifuged on a sucrose gradient with the same solutes. The gradient was divided into 17 fractions and portions were subjected to SDS-PAGE.

larger minor complex that may represent a dimer of the major complex (i.e. a ΔMBDΔPR dodecamer made by two incompletely cross-linked hexamers that became cross-linked to one another). I estimated the sedimentation coefficient of the major complex at 10S relative to tetrameric β-galactosidase (a 440 kDa tetramer), which would correspond approximately to a globular protein of 300 kDa. (ΔMBDΔPR is a 52 kDa protein.)

Gradient fractions that contained only the major complex, such as lane 9 from **Figure 3.12**, were pooled, concentrated, and subjected to β-mercaptoethanol at a range of concentrations to empirically determine the best concentration to produce a ladder when subjected to non-reducing SDS-PAGE (**Figure 3.13**). Once the optimal concentration of 0.25% β-mercaptoethanol was determined, the partially reduced complex was subjected to non-reducing SDS-PAGE on a 3-8% gradient Tris-acetate gel. These conditions yielded a ladder of six bands spaced at approximately 50 kDa intervals (**Figure 3.14**), of which the highest band is the same size as the non-reduced complex and the lowest is the correct molecular weight for a ΔMBDΔPR monomer (~52 kDa). I concluded that the major protein complex is a hexamer linked by p10-NTD disulfide bonds.

### Implications of the p10-CA interface for immature vs. mature NTD hexamer structure

The results above implied that the extended NTD structure is a dimeric version of the six-fold p10-CA hexamer observed in particles. This idea presented two difficulties. The first was that comparison of the amino acids involved in the RSV p10-CA interface with those involved in the only available hexameric NTD-NTD interface, that of MLV (87, 88), revealed a conflict: many

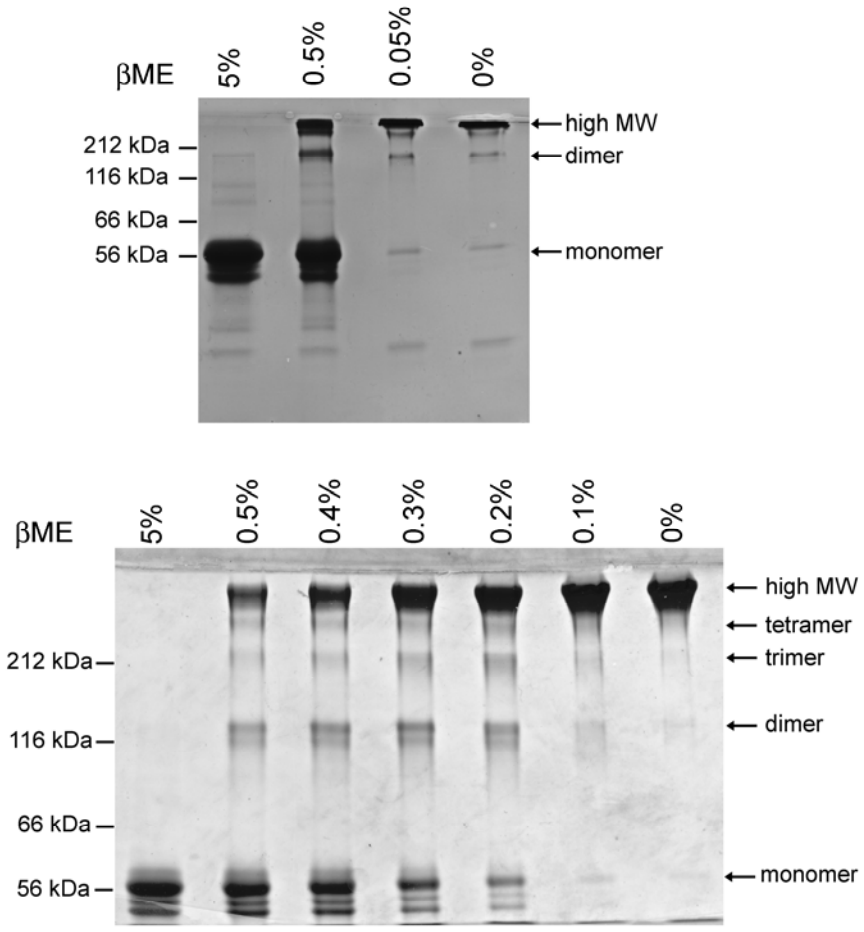



**Figure 3.13**: Optimal reducing agent concentration for partial reduction of the high MW complex was determined empirically. Gradient fractions containing purified "major" high MW complex, such as Figure 3.12 lane 9, were pooled, concentrated, mixed with indicated amounts of β-mercaptoethanol (βME) and analyzed by SDS-PAGE. Samples were boiled with SDS and the βME was added immediately before loading. **Top**: "Coarse" determination analyzed on 10% acrylamide gel. **Bottom**: "fine" determination analyzed on 7% acrylamide gel.

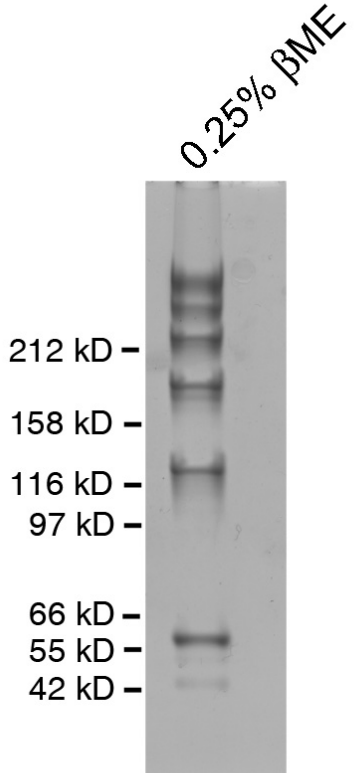



**Figure 3.14**: Major species in high MW Gag complex is a hexamer. Samples containing the major high MW band, such as lanes 9 and 10 in Figure 3.12, were pooled, concentrated, and partially reduced with β-mercaptoethanol. SDS-PAGE was performed on NuPAGE Tris-Acetate 3-8% gradient gels (Invitrogen).

structurally homologous residues are involved in both structures (**Figure 3.15**), implying that the extended NTD p10-CA interface and the hexameric CA-CA interface cannot exist simultaneously. This suggests that if a p10-CA interface exists in the immature NTD hexamer then the immature NTD hexamer must be significantly different from the mature NTD hexamer in order to accommodate p10. The second problem was that the unfolded β-hairpin region must bend to allow a next-neighbor p10-CA contact in the NTD hexamer. This is diagrammatically represented in **Figure 3.16**. The 25 residues of p10 are drawn as a hook that acts like a flap contacting a neighboring NTD. In the dimer the shaft of the hook is straight and the NTDs are symmetrically arranged, pointing in opposite directions, an orientation that is inferred to be biologically irrelevant. In the hexamer, the shaft of the hook must be bent to accommodate the same p10-NTD side chain contacts as in the dimer. To resolve these conundrums, I entered into collaboration with Paul S. Murray and Diana Murray, who have experience with molecular modeling, in order to model mature and immature RSV NTD hexamers.

The models generated by the Murray group are shown in **Figure 3.17**. In this panel the top three of the six subunits are drawn in space-filling style, with the protein-protein interactions shown by arrows between CA helices (numbers) or between the p10 segment (p) and CA helices. The bottom three subunits are drawn to show the polypeptide itself in ribbon style. The mature hexamer is based on the MLV CA NTD hexamer as determined by crystallography, which is stabilized primarily by interactions between the N-terminal β-hairpins of adjacent subunits and between the first helix of adjacent subunits, and secondarily by interactions between helix 2 and helix 3 of adjacent subunits (88). It should be noted that differences between the RSV

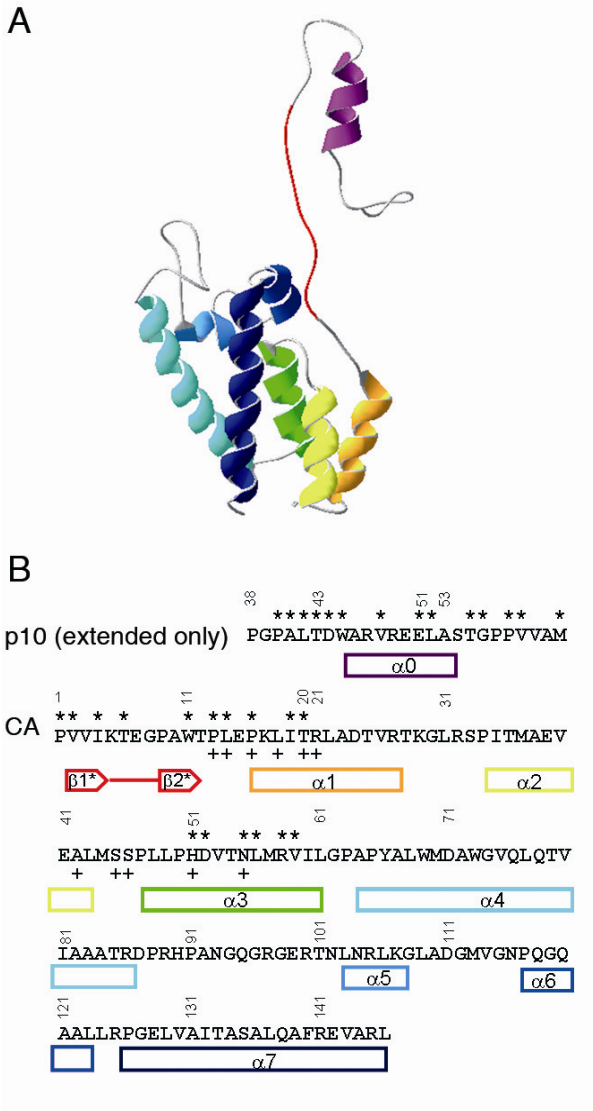



**Figure 3.15**: p10-CA interface from extended NTD structure overlaps with CA-CA interface in MLV NTD hexamer. **A**. Extended NTD monomer colored by helix. **B**. Annotated amino acid sequence of the RSV extended NTD, including 25 amino acids of p10. Asterisks above the sequence indicate residues involved in the extended NTD dimer interface (91). Plus signs below the sequence indicate the amino acids most structurally homologous to the MLV residues involved in the MLV mature NTD hexamer interface (88). Secondary structural elements are indicated with colors corresponding to panel **A** above; note that while the helices are valid for both structures, the β-hairpin (red) is present only in mature CA.

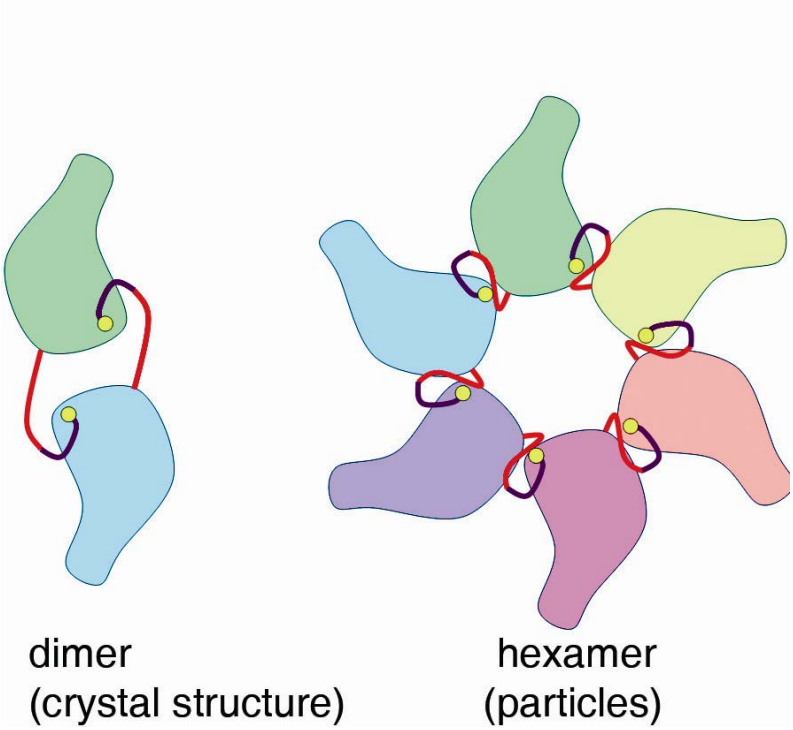



**Figure 3.16**: Schematic representation of p10-CA interaction in the extended NTD. **Left**: crystallographic dimer. **Right**: the hexamer proposed to form in immature particles. p10 is represented by the purple line, β-hairpin residues by the red line, and the remainder of the NTD by solid shapes. Small yellow circles represent the N-terminus of the extended NTD.

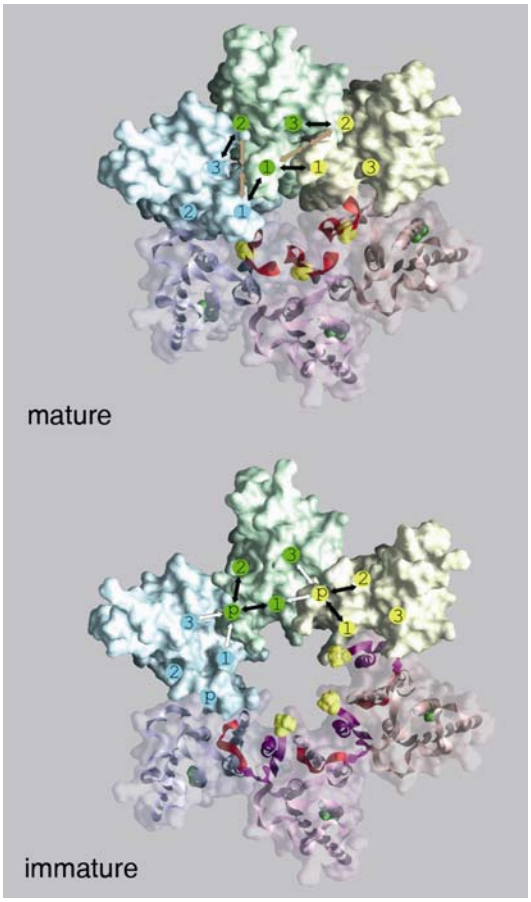



**Figure 3.17**: Molecular models of mature and immature RSV CA NTD hexamers with approximate helix positions and interactions indicated. **Top**: Molecular model of mature RSV NTD hexamer based on mature MLV NTD structure. The upper three subunits of the hexamer are shown as solid space-filled surfaces while the lower three subunits are shown as transparent surfaces superimposed on ribbon diagrams. On the ribbon diagrams the N-terminus is space-filled in yellow and the C-terminus in green. The β-hairpin ribbon is in dark red. Approximate helix positions are indicated by labeled circles. Dark arrows indicate direct intermolecular CA-CA interactions predicted from the MLV NTD hexamer and tan arrows indicate predicted water-mediated interactions. **Bottom**: Molecular model of immature RSV NTD hexamer based on extended NTD structure and mature MLV NTD hexamer. On the ribbon diagrams p10 residues are dark purple and the unfolded β-hairpin residues dark red; the depiction is otherwise similar to the mature model. White arrows indicate intermolecular p10-CA interactions and black arrows indicate possible intramolecular interactions.

and MLV monomer structures mean that the relative helix positions within the hexamer are somewhat different as well, and thus the interactions inferred from the MLV structure must be considered approximate for RSV.

The addition of p10 in the immature model expands the hexamer (**Figure 3.17**, bottom), opening up a much larger hole in the center. While the nature of the model does not allow for exact measurements, the immature hole is approximately three to four times larger than the mature hole (88), approximately 42 Å compared to approximately 12 Å for the mature hexamer. A similarly large hole in the NTD hexamer of the immature HIV-1 Gag lattice was recently observed in immature HIV-1 particles analyzed by electron cryo-tomography (130). I provided a sample of *in vitro* assembled RSV VLPs for analysis by the same technique, and the RSV particles also have large central cavities in the NTD hexamers (John Briggs, personal communication). The cryotomography results are consistent with the Murray model for the immature RSV NTD hexamer, providing another line of evidence that the p10-CA interface depicted in the extended NTD structure is present in the immature NTD hexamer.

## DISCUSSION

Retroviral assembly is a multi-step process in which the structural protein Gag assembles into immature particles, and then is cleaved by the viral protease PR into products that re-assemble to form the internal structure of the mature infectious virus. The CA domain of Gag forms a hexameric lattice in both immature and mature particles, but differences in the lattice suggest that the Gag-embedded CA and the liberated CA make different

protein-protein contacts. The nature of these differences is not well understood. The N-terminal β-hairpin of free CA is highly conserved and is important for correct mature assembly of at least some (123) retroviruses. In addition, Rous Sarcoma virus has a specific sequence requirement for correct immature assembly, namely the C-terminal 25 amino acids of the p10 domain that adjoin CA (58). A high-resolution structure of an RSV NTD extended to include this critical region of p10 showed a dimer with an extensive p10-NTD interface (91), but the relevance of this structure to immature assembly was uncertain prior to the work presented in this thesis.

I have now provided evidence that the p10-NTD interface identified in the structure is relevant to immature assembly, but that in immature virus particles the interface forms a hexamer rather than a dimer. Since the NTD is believed to form a hexamer in both the immature and the mature CA lattices, I concluded that p10 participates in the immature NTD hexamer. This is the first example of a critical heterologous interdomain Gag-Gag contact in retroviral assembly.

The interaction between p10 and the NTD in the extended NTD structure is incompatible with the NTD-NTD interactions seen in the MLV mature NTD hexamer structure (88). Molecular modeling of an immature RSV NTD hexamer, based on the RSV extended NTD dimer and the MLV hexamer structures, shows that the p10-NTD interaction forces the NTD subunits further apart, opening a central cavity. The final NTD arrangement is strikingly similar to that seen for immature HIV-1 particles (130) and RSV particles (John Briggs, personal communication) examined by electron cryotomography. This similarity suggests that immature retroviral lattices are structurally homologous, just as are their mature CA proteins. However, it is unknown by

what mechanism the HIV-1 NTD lattice is expanded in the immature lattice as compared with the mature lattice.

*In vitro* studies of HIV-1 assembly show that assembly pH acts as a morphological switch for immature-like assembly: an HIV-1 Gag protein lacking the C-terminal p6 domain and amino acids 16-99 of MA assembles only into spheres at pH 8.0 but assembles into both spheres and tubes at pH 6.5 (50). The assembly morphology switch is paralleled by changes in epitope availability in the NTD of monomeric Gag, with epitopes in helix 3 and helix 6 available at pH 8.0 but masked at pH 6.5 (50). These results suggest that for HIV-1 a pH-mediated conformational change in Gag – which may reprise changes caused by PR cleavage *in vivo* – is responsible for the switch between immature-like spheres and mature-like tubes. The molecular nature of this switch remains to be unraveled.

Electron cryotomography has recently provided the first data on the retroviral immature lattice (130). This study shows that in contrast to the mature lattice, the immature lattice has the NTD hexamers more widely spaced, with almost no NTD-NTD contact within the hexamer, and the CTDs more tightly packed. The hexamer appears to be stabilized by interactions among the spacer domains C-terminal to the CTD, a region that is also important for RSV assembly (60) In addition, a novel “domain-swapped” structure has been proposed for the HIV-1 CTD (54); this dimer is much tighter than the traditional dimer structure seen for both purified protein (42) and the reconstructed mature lattice (43), and it is tempting to speculate that the immature lattice, with its greater stability and tighter CTD spacing, incorporates the domain-swapped CTD dimer instead.

The model of the RSV immature NTD hexamer presented above is similar to the NTD hexamers observed by cryotomography, with widely spaced monomers and a large central cavity. However, in contrast to the immature HIV-1 lattice, in which neither the NTD nor MA appears to stabilize the hexamer (130), the RSV immature hexamer is stabilized, at least in part, by p10-NTD interactions. This difference is not surprising: previous studies with chimeric RSV-HIV-1 Gag constructs have shown that constructs with RSV CA require the critical p10 sequence from RSV upstream for assembly, while constructs with HIV-1 CA tolerate RSV sequence upstream (2), suggesting that the HIV-1 CA domain does not interact with upstream Gag sequences. Similar studies with chimeric MLV and spleen necrosis virus (SNV) Gag constructs show that assembly of these distantly related gammaretroviruses also requires that the domain upstream of CA (p12) be homologous to the CA domain. I therefore speculate that the gammaretrovirus immature hexamer may be stabilized by interactions between CA and p12.

The incompatibility of the immature p10-CA interaction with the mature MLV hexamer supports the idea that retroviruses have distinct immature and mature NTD-NTD interfaces, with PR cleavage favoring the mature over the immature interface. The participation of the upstream p10 domain in the immature lattice is intriguing because it suggests a mechanism by which proteolysis leads to maturation. The RSV p10 domain both stabilizes the immature hexamer and prevents mature NTD-NTD contacts from forming. The maturation process can thus be imagined as follows: upon PR activation, cleavage between p10 and CA destabilizes the immature hexamer and allows p10 to diffuse away from CA, which in turn unmasks mature contact surfaces and allows the mature NTD hexamer to form (**Figure 3.18**).

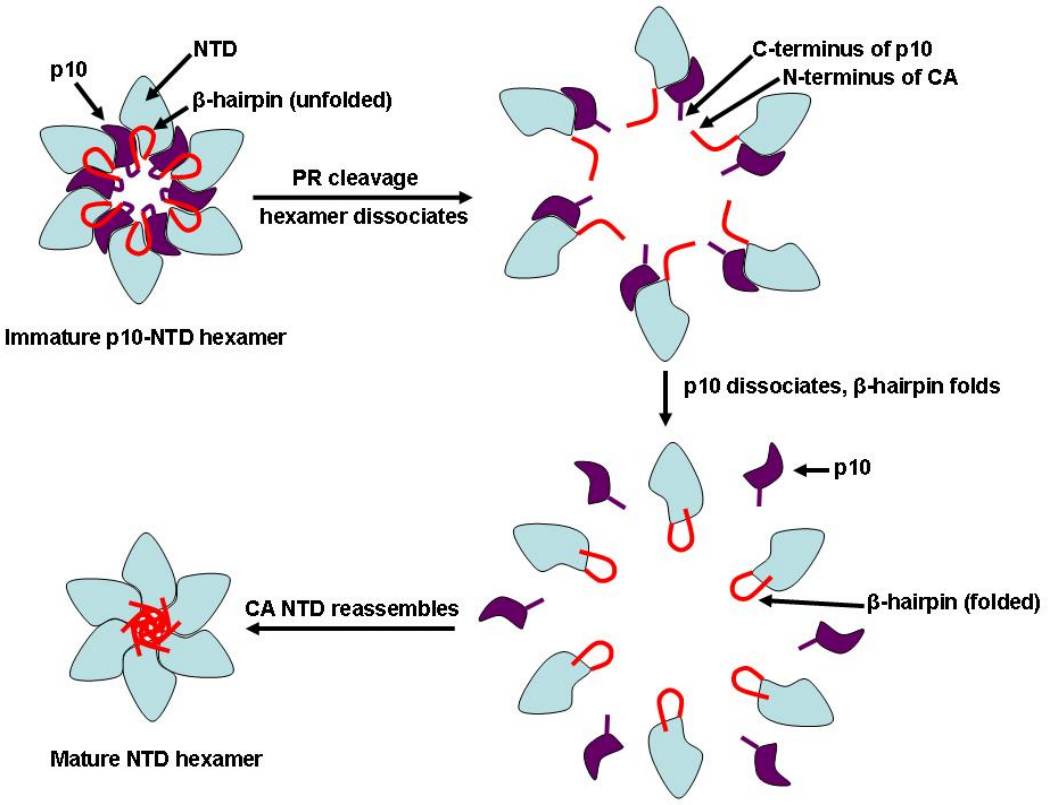



**Figure 3.18**: RSV CA NTD hexamer maturation model. PR cleavage between p10 and CA allows the immature hexamer to dissociate. The p10 domain diffuses away, allowing the β-hairpin to fold. The mature NTD then reassembles into the mature hexamer.

The nature of the link that couples proteolysis to maturation is likely to differ among retroviral families. There are three broad categories of mechanisms by which PR cleavage can change the CA-CA contacts: by initiating refolding of CA, by removing sequences that promote immature contacts, or by removing sequences that inhibit mature contacts. The first mechanism, exemplified by the refolding of the β-hairpin, is well documented.

Disrupting the β-hairpin of Mason-Pfizer monkey virus (MPMV) CANC by removing the N-terminal proline converts the *in vitro* assembly products from disorganized aggregates into spherical particles (120), suggesting that proteolytic refolding is incompatible with the immature lattice. HIV-1 appears to use both refolding (based on pH-dependent epitope availability) and removing required sequences; since the immature hexamer is stabilized by interactions between downstream SP1 domains, proteolysis should dissociate the immature hexamer. For RSV, the contact between p10 and CA both stabilizes the putative immature NTD hexamer and interferes with the inferred mature NTD-NTD contacts, so PR cleavage between p10 and CA both destabilizes the immature particle and promotes mature assembly. Thus, while the structural properties of both immature and mature retroviral particles appear to be similar across families, the mechanics of maturation may prove to be quite different.

# CHAPTER FOUR

# ASSEMBLY OF ROUS SARCOMA VIRUS GAG BY ARTIFICIAL DIMERIZATION

## INTRODUCTION

Dimerization appears to be a critical initial step in retroviral assembly. HIV-1 Gag has been demonstrated to bud from cells when the NC domain of is replaced by a dimeric “leucine zipper” (1, 137). The same result is seen for RSV in a baculovirus/Sf9 expression system, with normal particle morphology confirmed by electron microscopy (56); this result had not, at the inception of this thesis research, been confirmed in native (chicken) cells. Dimerization seems sufficient to induce assembly *in vitro* as well: while diverse nucleic acids support assembly, there is a minimum oligonucleotide size that for RSV corresponds to two NC-binding sites (79). In addition, both dimerizing and trimerizing leucine zippers have been used to drive HIV-1 assembly *in vitro* (Alan Rein, personal communication). Replacement of HIV-1 NC with a single cysteine residue that is oxidized to form disulfide-mediated dimers also leads to assembly *in vitro* (4). In the case of RSV, for which the CTD of CA does not dimerize in solution as it does for HIV-1, C-terminal Gag dimerization has been hypothesized to cause a conformational change in the CTD of CA, activating the CTD for Gag multimerization and assembly (78). While a mathematical argument suggests that dimerization will increase the probability of assembly even in the absence of any such activation step (140), ample evidence indicates that dimerization plays a specific role in retroviral assembly.

No artificial dimerization system has proven successful for RSV assembly *in vitro*. Attempts by a former student in this laboratory to replace the NC domain with a leucine zipper in the RSV *in vitro* system yielded a poorly soluble protein that failed to assemble *in vitro*; thin section TEM of *E. coli* expressing this construct showed only low numbers of excessively small (25-30 nm) particles (77). One hypothesis to explain this failure is that in this construct, a functional element that spans the SP-NC cleavage site was interrupted when NC was replaced by a dimerization domain (**Figure 4.1**). This phenomenon was first seen when RSV NC was replaced with a leucine zipper in the baculovirus expression system. While the simple fusion produced correct spherical particles, inclusion of a five amino acid spacer between SP and the zipper produced cylindrical particles (56) (**Figure 4.2**). Further mapping of the element showed that it spans the C-terminal tail of CA, the SP domain, and the first four amino acids of NC (60) (**Figure 4.1**). Disruption of this sequence by insertion or substitution produced cylindrical Gag particles in insect or chicken cells (60); constructs parallel to those that make cylindrical particles in cells failed to assemble *in vitro* (59). The role of this element in RSV assembly is not fully characterized, although it likely relates to the internal pillar-like structures seen in cryo-tomograms of HIV-1 (130) and RSV (John Briggs, personal communication) immature particles and described in Chapter One of this thesis (**Figure 1.10**, magenta structures).

The previous leucine zipper chimera involved fusing the zipper domain directly to the C-terminus of SP (**Figure 4.1**), thus removing the last four amino acids of the CA-SP-NC element. While such a fusion formed spherical particles in the baculovirus expression system, it is possible that the degree of overexpression inherent to that system masks some subtle defect in

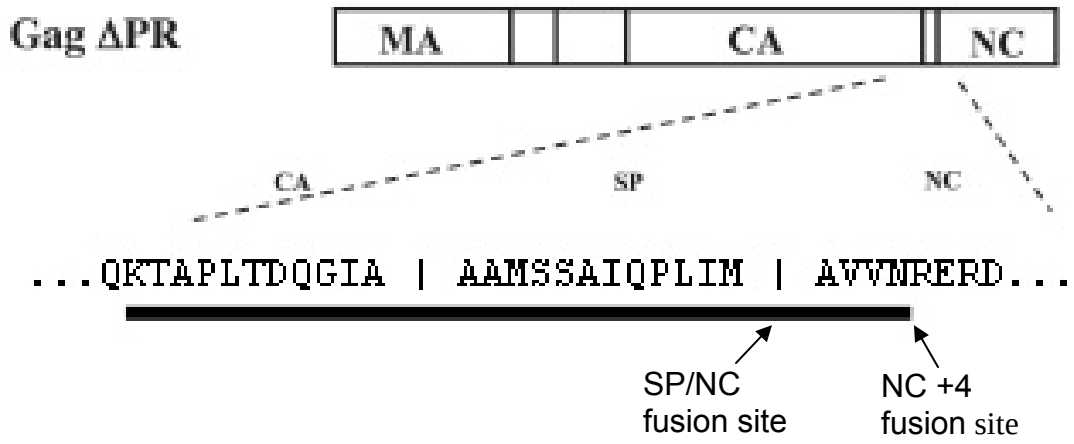



**Figure 4.1**: Novel morphology-determining element in RSV Gag spans the CA-SP-NC cleavage sites. CA-SP-NC element is indicated by heavy black line. PR cleavage sites are indicated by vertical lines. Sites of Gag-dimerization domain fusions are indicated by arrows. Figure adapted from reference (60).

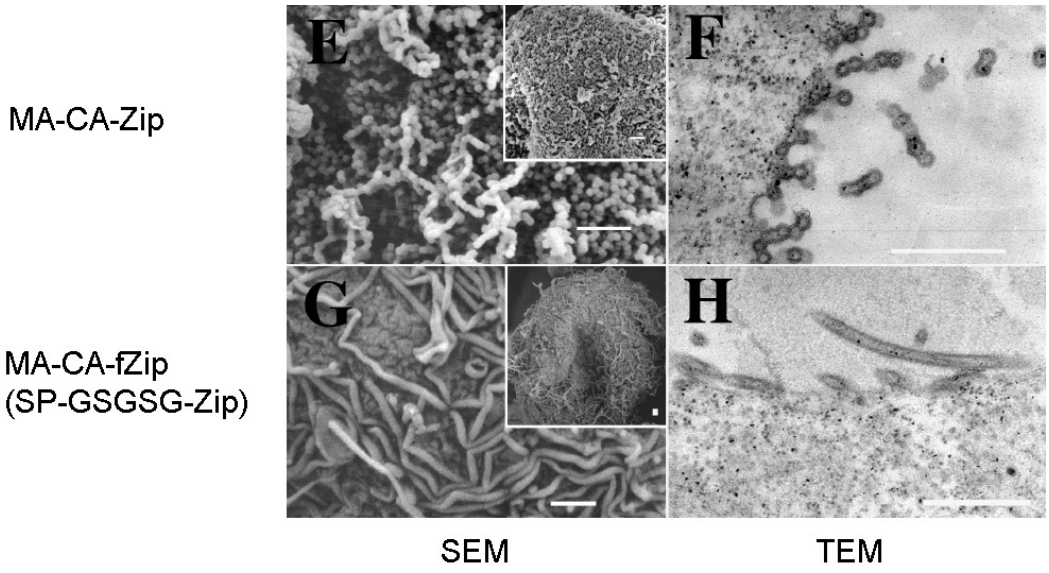



**Figure 4.2**: Replacement of RSV NC with leucine zipper domain in the baculovirus expression system yields morphologically normal particles, but inclusion of a five amino acid linker (GSGSG) between SP and the zipper yields cylindrical particles. Figure taken from reference (56).

assembly. Inclusion of the entire CA-SP-NC element may allow for an efficient *in vitro* leucine zipper-mediated assembly system for RSV.

Artificial dimerization-assembly systems are of interest in part because of the additional features that they contribute to assembly studies. A feature of particular interest is the potential to make assembly inducible. In cells, retroviral assembly proceeds automatically when a sufficient level of Gag expression is reached. The lack of control post-expression imposes a time lag between the time that expression is induced and the time that assembly can be detected. For kinetic studies, it would be useful instead to be able to induce assembly of a pre-formed pool of Gag protein. One potential candidate for an inducible system of retroviral assembly is the Argent Regulated Homodimerization System developed by Ariad Pharmaceuticals. This system is based on the interaction of human FK506 binding protein (FKBP) and its ligand, the small molecule FK506, a natural product that acts as an immunosuppressive drug. Homodimerization of FKBP is induced with homobifunctional ligands in which two molecules of FK506 are tethered. These ligands penetrate the plasma membrane and are thus readily introduced into cells. Ariad has developed modified versions of FKBP (the mutant used in these studies is designated $F_v1$) and FK506 (AP20187 was used in these studies) which are specific to each other and do not interact with cellular FKBP domains (23). Therefore, fusion of a protein of interest to a modified FKBP domain, followed by introduction of the ligand, is expected to dimerize the protein of interest rapidly and specifically (**Figure 4.3**). Replacement of the NC domain of Gag with FKBP might allow for rapid, specific induction of retroviral assembly in cells or *in vitro*, which in turn could facilitate studies of retroviral assembly kinetics.

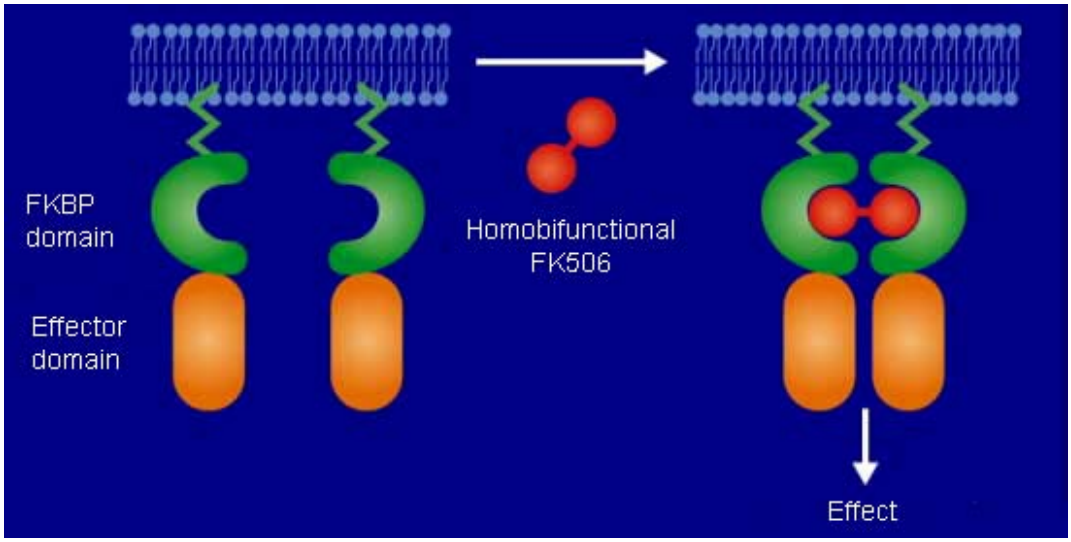



**Figure 4.3**: Principle of FKBP-FK506 inducible homodimerization system. The FKBP domain is fused to an effector domain. The homobifunctional FK506 ligand binds to two FKBP domains, forcing the effector domains to dimerize and potentiating an effect. Figure adapted from documentation for ARGENT™ Regulated Homodimerization Kit, Version 2.0, from ARIAD Pharmaceuticals, Inc.

## RESULTS

### Gag-leucine zipper fusion does not bud consistently from avian cells

The Gag-leucine zipper fusion in which the leucine zipper domain replaces NC at the PR cleavage site assembles correctly in the baculovirus expression system but not *in vitro*. To clarify whether this construct is truly assembly competent, I moved the MA-CA-LZ, MA-CA-fLZ, and MA-CA constructs from the baculovirus study (56) into a CMV-driven vector (**Figure 4.4**) and expressed these constructs in DF1 (embryonic chick fibroblast) cells by transient transfection. I used the previously described Gag-D37S construct (69), which was originally made in the same CMV-driven vector plasmid, as a positive control for budding. Budding was assayed by $^{35}$S-methionine labeling, SDS-PAGE, and autoradiography.

Initial results of these experiments were promising. Both the MA-CA-LZ and MA-CA-fLZ constructs were released from cells at readily-detectable levels (**Figure 4.5**). Consistent with previous results (56), the MA-CA construct, which lacks either NC or a dimerization domain, was not released, indicating that NC plays a role in assembly that can be replaced by the leucine zipper domain. Unfortunately, these results could not be reproduced consistently, and many transfections resulted in little to no release of the leucine zipper constructs (**Figure 4.6,** compare MA-CA-LZ “medium” lanes after 8 hours, top and bottom). Budding efficiency appeared to be related to the level of protein expression (**Figure 4.6**, compare MA-CA-LZ “lysate” lanes after 1 hour, top and bottom), but experiments to confirm this observation (for example, varying the amount of plasmid transfected into cells and comparing amount of DNA transfected to budding efficiency) were not performed. While

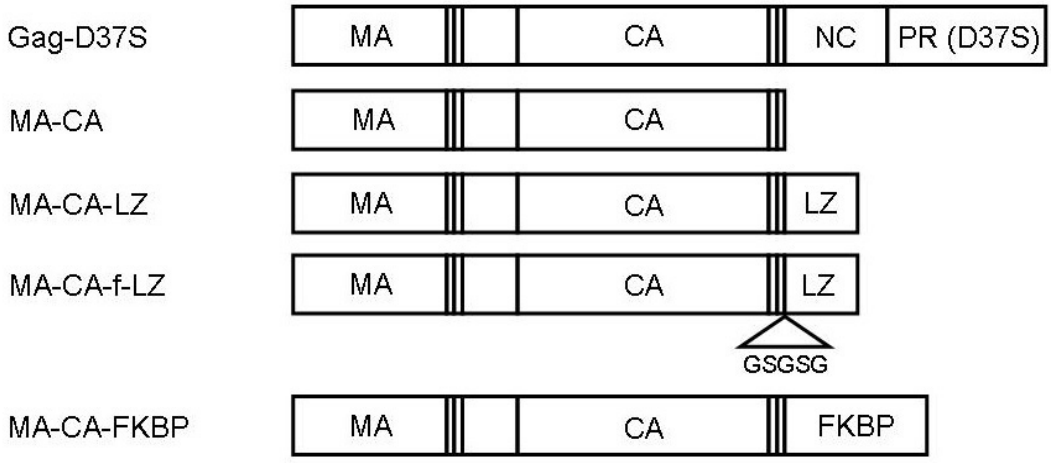



**Figure 4.4**: CMV-driven Gag constructs used for Gag dimerization studies in DF1 cells. D37S is an active site mutation that inactivates RSV PR.

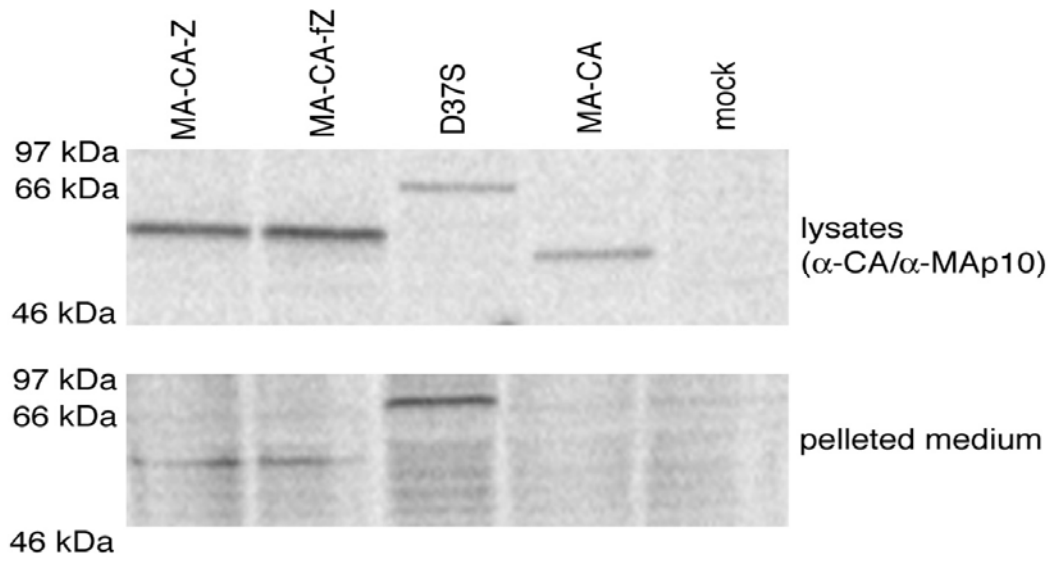



**Figure 4.5**: Gag-leucine zipper fusion proteins bud from DF1 cells. Cells were transfected with indicated constructs, pulse-labeled with $^{35}S$ methionine 24 hours post-transfection, and analyzed after a one hour chase. **Top**: Gag proteins immunoprecipitated from cell lysates using a mixture of α-CA and α-MA-p10 antibodies. **Bottom**: Gag particles pelleted from cell culture medium. Note that MA-CA does not bud.

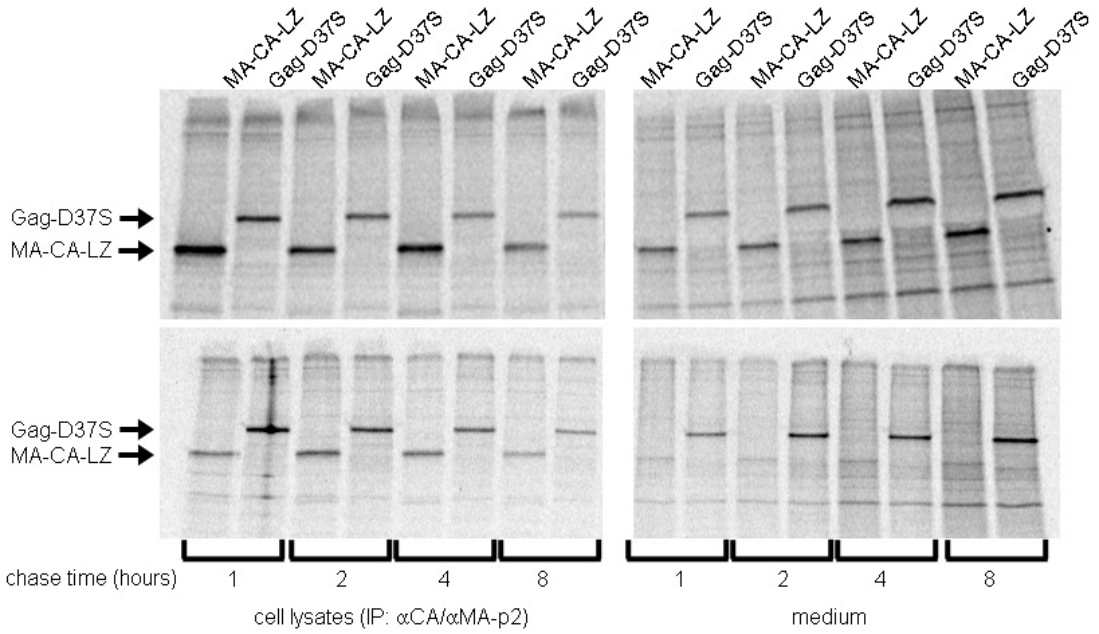



**Figure 4.6**: Budding of Gag-leucine zipper fusion proteins was inconsistent. Two pulse-chase time courses are shown for comparison. Cells were pulse-labeled 24 hours post-transfection and analyzed after the chase time indicated. **Top**: Successful budding of MA-CA-LZ. Note strong appearance of lower Gag band in medium (right). **Bottom**: Poor budding of MA-CA-LZ. Note relative absence of lower Gag band in medium (right).

the budding rate for MA-CA-LZ was, at its best, comparable to that for Gag-D37S (**Figure 4.7**), I was unable to obtain statistically useful data due to the inconsistency of budding. In light of the inconsistency, I judged thin section TEM to examine the particle morphology to be too time-consuming and did not attempt to perform it. My tentative conclusion based on these findings is that the substitution of the leucine zipper for the NC domain at the SP-NC cleavage site introduces a subtle assembly and/or budding defect that permits budding only under certain optimal conditions.

**Gag-leucine zipper fusions at the NC+4 position do not assemble *in vitro***

The failure to assemble a Gag-leucine zipper construct *in vitro* had been ascribed to the previously-unsuspected existence of a functional element including the first four amino acids of NC. To address this hypothesis, I generated a construct from ΔMBDΔPR in which the CREB bzip domain replaced NC at the NC+4 position. A five amino acid flexible linker (GSGSG) was included in order to minimize interference between the two domains (**Figure 4.8**). This construct was readily expressed in and purified from *E. coli*. Unfortunately, the construct failed to produce particles under standard *in vitro* assembly conditions (**Figure 4.9**). Furthermore, rate-zonal sedimentation analysis of ΔMBD-bzip did not show a significant fraction shift relative to monomeric ΔMBDΔPR, suggesting that the protein did not dimerize (**Figure 4.10**) A second ΔMBD-bzip gradient was more ambiguous (data not shown), admitting the possibility that some degree of dimerization may have occurred in that experiment, but overall the evidence for dimerization of ΔMBD-bzip was not convincing.

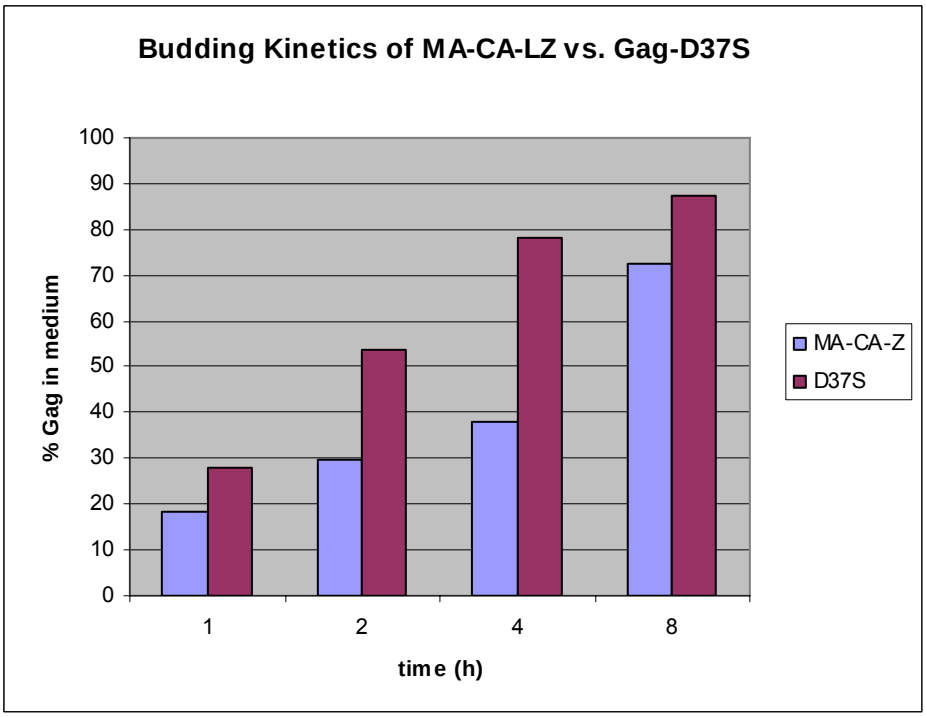



**Figure 4.7**: Budding rate of MA-CA-LZ is comparable to that of Gag-D37S. Budding is expressed as percentage of total Gag found in the medium at each time point.

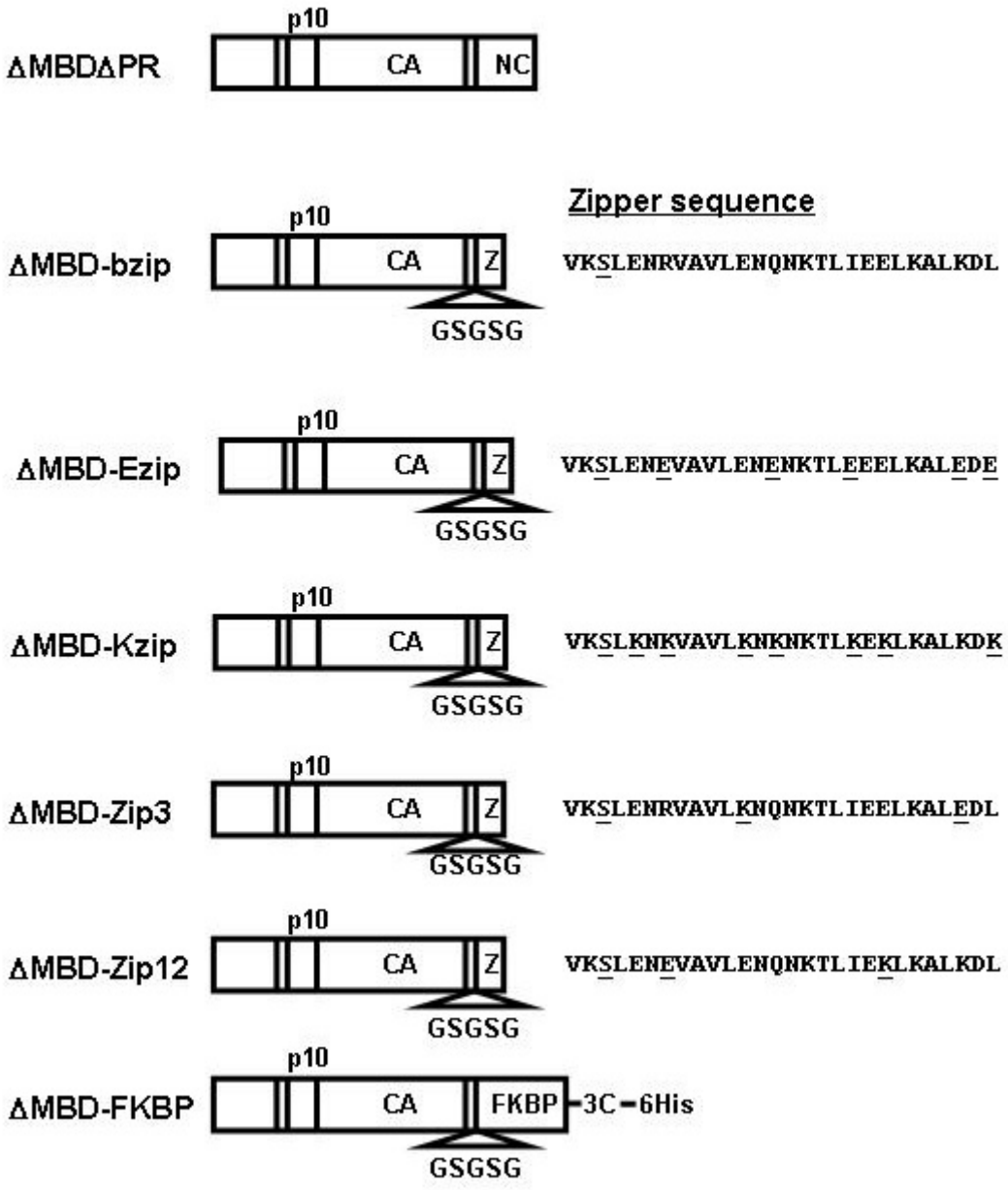



**Figure 4.8**: Constructs used for *in vitro* dimerization/assembly assays. Mutated amino acid residues are underlined.

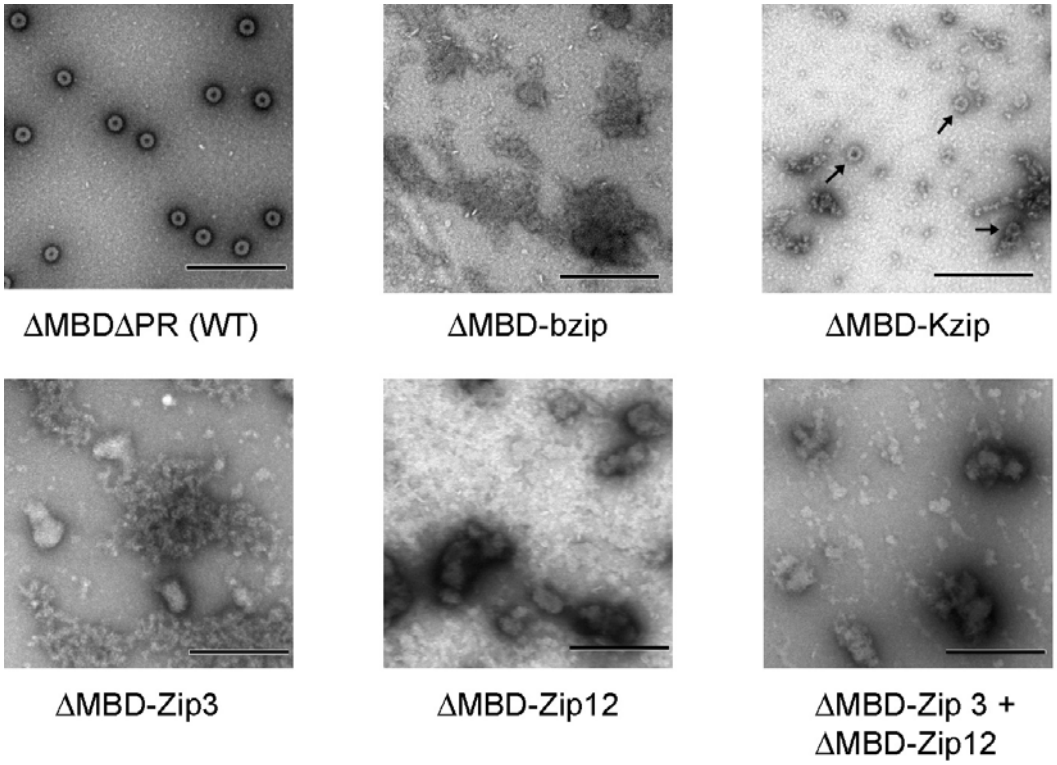



**Figure 4.9**: Negative stained TEM images of wild-type and leucine zipper fusion proteins assembled *in vitro*. Scale bars represent 100 nm. Arrows denote VLP-like structures seen for ΔMBD-Kzip.

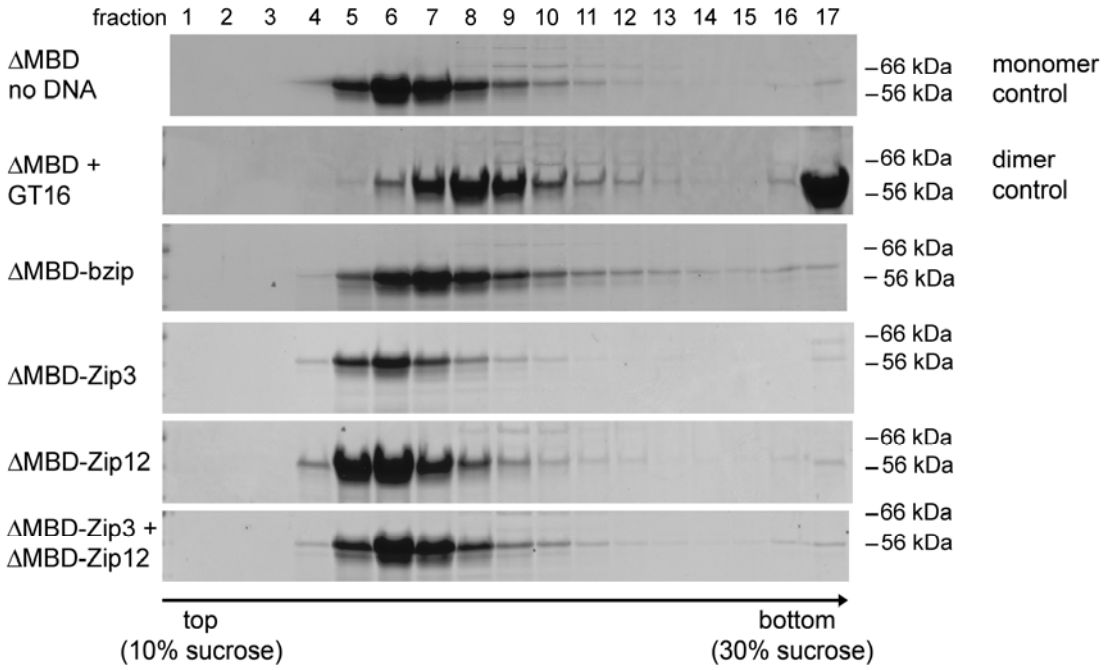



**Figure 4.10**: Rate-zonal sedimentation analysis of leucine zipper fusion dimerization. Samples were loaded onto 10-30% sucrose gradients and centrifuged for 24 hours. Arrow indicates direction of sedimentation. ΔMBD (no DNA) is a monomeric control and ΔMBD + GT16 is a dimeric control. None of the zipper proteins appears to dimerize.

It is possible that the protein fails to dimerize or assemble because dimerization-competent protein is lost during the purification process. This possibility was addressed by creating two pairs of heterodimeric leucine zippers. The first pair, designated Ezip/Kzip (**Figure 4.8**), had previously been used to assemble HIV-1 Gag in cells (137). All eight of the amino acids in the "e" and "g" positions flanking the hydrophobic interface were changed to either glutamic acid or lysine. These changes effectively prevent the formation of homodimers due to electrostatic repulsion but favor the formation of heterodimers (**Figure 4.11**). Neither protein is expected to dimerize or assemble when expressed alone, but dimerization and possibly assembly should occur when the two purified proteins are mixed.

In practice, the Ezip fusion construct could not be expressed in *E. coli*, and the arrest in *E. coli* growth that occurred upon induction suggested that the protein was toxic (data not shown). The Kzip construct was efficiently expressed and purified using the standard protocol for ΔMBDΔPR. As Kzip carries a net positive charge, I hypothesized that it might function as a replacement for NC by electrostatic binding to nucleic acid. A single assembly trial was made by mixing the Kzip protein with 10% ssDNA under standard *in vitro* assembly conditions and examining by negative staining and TEM (**Figure 4.9**). While some of the protein debris was suggestive of assembly (**Figure 4.9**, arrows), no definite VLPs were observed. Nucleic acid-binding was not assayed and no attempt was made to determine whether the ratio of protein to nucleic acid was appropriate for this construct; therefore, it remains possible that ΔMBD-Kzip could assemble on nucleic acid if assembly conditions were optimized.

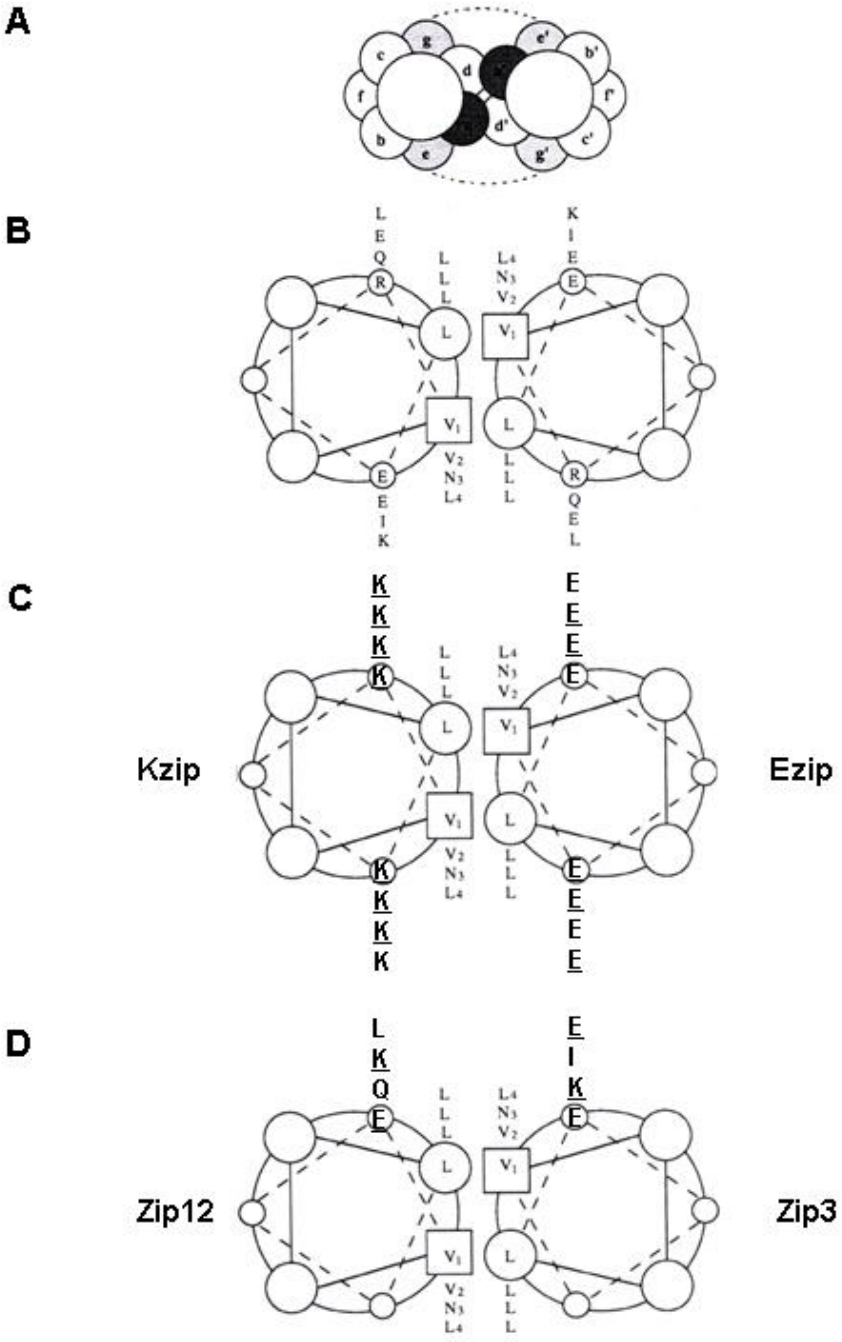



**Figure 4.11**: Helical wheel representations of heterodimeric leucine zipper constructs. Mutated amino acid residues are underlined. Figure adapted from reference (76).

A final attempt at leucine zipper-mediated heterodimerization was made using the previously published heterodimeric zippers Zip3 and Zip12 (76). The investigators forced the proteins to heterodimerize by introducing electrostatic clashes at two e or g positions in each homodimer (**Figure 4.11**). In the original study, the a3 asparagine of Zip12 was mutated to histidine to prevent Zip12 from forming any homodimers. As including the histidine at a3 in both zippers prevented heterodimerization as well, the a3 asparagine of Zip3 was left in place. Since I did not prefer either possible homodimer, I restored the a3 asparagine of Zip12 and changed only the charged residues in the e and g positions. Both ΔMBD-Zip3 and ΔMBD-Zip12 (**Figure 4.8**) were successfully expressed in and purified from *E. coli*, but they did not form VLPs either alone or when mixed (**Figure 4.9**). Rate-zonal sedimentation analysis showed that both proteins were monomeric in solution, but there was no evidence that the two proteins were able to form heterodimers when mixed (**Figure 4.10**).

In conclusion, as of this writing I have not been able to demonstrate assembly of RSV Gag by leucine zipper-mediated dimerization.

**Gag-FKBP fusion proteins fail to respond to dimerization reagent**

In order to investigate the possibility of an inducible Gag dimerization assembly system in cells, I generated a CMV-driven construct in which the modified FKBP domain $F_v1$ (which has mutations conferring altered ligand-binding specificity (23)) replaced the NC domain of Gag at the SP-NC cleavage site. (This work was begun before the characterization of the CA-SP-NC element.) The PR domain of Gag was not included. The resulting constructs were transfected into DF1 cells. After 24 hours, cells were treated with AP20187 (a homobifunctional FK506 analog that is bound by $F_v1$ but not

by cellular FKBP domains) at a range of concentrations. Cells and medium were harvested 24 hours after AP20187 treatment and analyzed by immunoblotting. Surprisingly, while the Gag-FKBP fusion did appear to bud from cells, the amount of Gag seen in the medium was approximately the same at every concentration of AP20187 used, including mock-treated cells that were not exposed to AP20187 at all (**Figure 4.12**). As budding was independent of AP20187, the FKBP fusion did not appear to function as predicted, I discontinued this line of investigation and did not perform thin section TEM to verify particle morphology.

I made a second attempt to assemble RSV Gag using inducible FKBP dimerization, this time in the *in vitro* assembly system. I generated a construct, ΔMBD-FKBP (**Figure 4.8**), in which the NC domain of the ΔMBDΔPR protein was replaced with FKBP at the NC+4 position (**Figure 4.1**). A five amino acid GSGSG linker was included at the Gag-FKBP junction. The FKBP domain included a C-terminal 6His tag separated from the FKBP sequence by a 3C protease cleavage site to allow for later removal of the 6His tag if desired. The fusion protein was expressed in *E. coli* and purified by Ni-NTA agarose chromatography.

The ΔMBD-FKBP fusion protein was adjusted to standard *in vitro* assembly conditions and a stoichiometric concentration of 50% (10 μM) of AP20187 was added. The reaction was incubated under standard *in vitro* assembly conditions and assayed by negative staining and TEM. No virus-like particles were observed (data not shown).

I then assayed whether the chimeric ΔMBD-FKBP protein was able to dimerize by diluting the protein to the *in vitro* assembly concentration (20 μM)

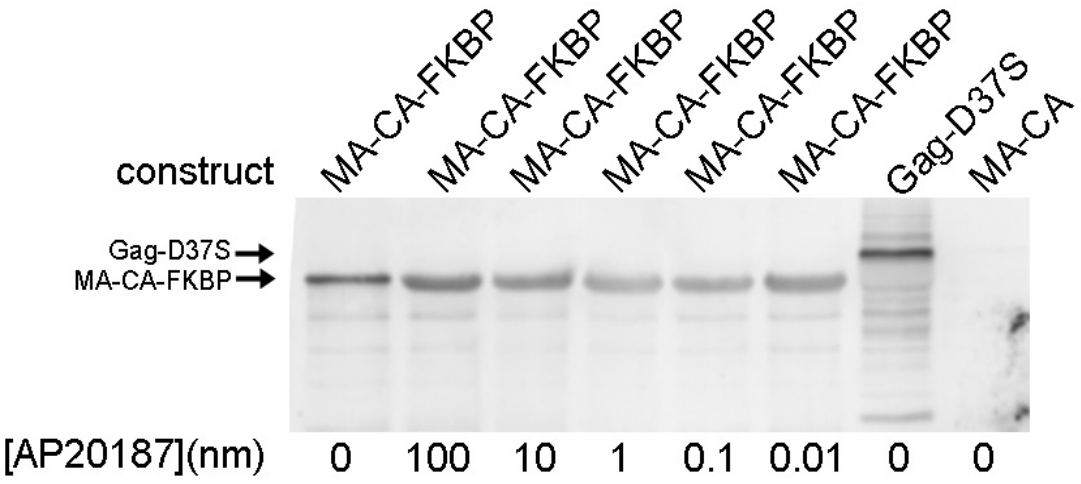



**Figure 4.12**: MA-CA-FKBP release from DF1 cells is independent of dimerizing ligand AP20187. Cells were treated with AP20187 24 hours post-transfection. Medium from cells was collected 24 hours post-treatment and particles were collected as described in the Materials and Methods. Pellets were analyzed by SDS-PAGE and immunoblotting using α-CA antibody. Gag-D37S was used as a positive control for budding and MA-CA as a negative control.

at pH 8.0, adding a stoichiometric amount (10 μM) of AP20187, and subjecting the mixture to rate-zonal sedimentation. No peak shift was observed in response to AP20187 (**Figure 4.13**), suggesting that the ΔMBD-FKBP fusion did not dimerize in response to AP20187.

## DISCUSSION

While the overall failure to produce RSV Gag assembly by artificial dimerization is disappointing, several aspects of these studies remain open for investigation. The success of leucine zipper-mediated assembly for HIV-1 and for RSV in the baculovirus system, and the partial success reported here for RSV in DF1 cells, all suggest that this approach might be viable under some as-yet unidentified set of conditions.

One possibility for the assembly failure of the Gag-leucine zipper fusion proteins presented here is that the CA-SP-NC element remains disrupted. The C-terminal extent of the element was defined by insertion of the GSGSG linker at various positions in NC (60). This approach leaves the remainder of the NC domain intact and is thus not entirely analogous to replacing the downstream portion with a completely different sequence. In fact, the constructs which couple the zipper domain directly to the end of SP and omit the last 4 amino acids of NC, such as the CMV-driven Gag-LZ used in this study and the MA-CA-LZ construct used in the baculovirus study, have generally been more successful than those incorporating a linker between the Gag and zipper sequences. This trend, combined with the hypothesis that the CA-SP-NC element forms a long α-helix, suggests that the linker sequence itself may be detrimental to assembly, possibly by breaking a helix. The downstream portion

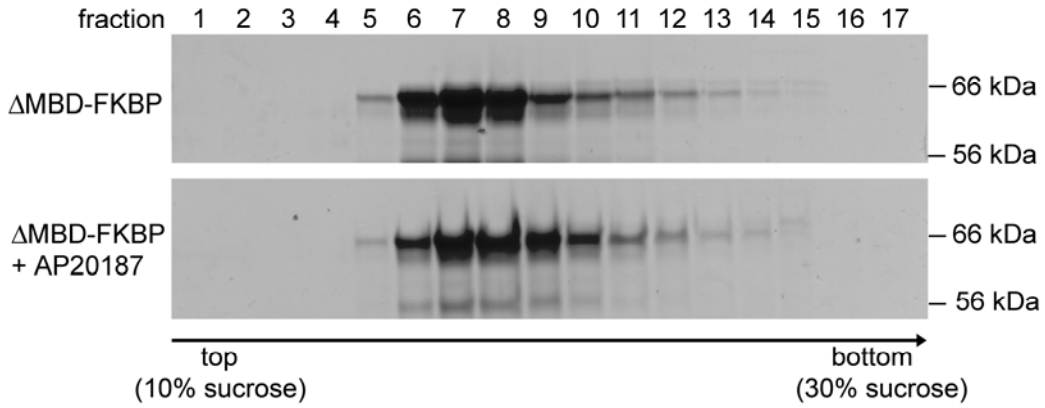



**Figure 4.13**: Rate-zonal sedimentation analysis of ΔMBD-FKBP dimerization in response to AP20187. Samples were loaded onto 10-30% sucrose gradients and centrifuged for 24 hours. The arrow represents the direction of sedimentation. The protein does not appear to dimerize in response to the drug.

of NC may be able to compensate for this putative helix-breaking in a way that the zipper domain cannot. It may also be noteworthy that the construct in which GSGSG was inserted at the NC+4 position, which appeared to produce normal particles in cells, assembled relatively poorly *in vitro* (59), leaving open the possibility of a subtle assembly defect that was not apparent in the cell-based studies. Future studies in which the linker was eliminated and the position of the zipper fusion extended C-terminally might produce more favorable results, although the latter suggestion requires including basic residues of NC that could confound interpretation either by recruiting nucleic acid or, in the absence of nucleic acid, by interfering with Gag-Gag interactions due to electrostatic repulsion.

The failure of the purified Gag-zipper fusions even to dimerize *in vitro* suggests three possibilities. The first is that dimerization did in fact occur, as suggested by the single ambiguous ΔMBD-bzip gradient (data not shown), but that the method used to detect dimerization, rate-zonal sedimentation, was not sufficient to the task. It is true that small shifts of one or two gradient fractions may be masked by technical difficulties inherent in fractionating the gradients. It is also possible that the dilution that inevitably occurs as the protein moves down the sucrose gradient led to dissociation of any dimers that formed. Further experiments using more definitive techniques, size-exclusion chromatography for example, would be required to say definitively that the proteins do not dimerize *in vitro*.

The second possibility is that mutation of the cysteine at the beginning of the zipper, intended to simplify future cross-linking studies, may have interfered with dimerization. I do not consider this probable, as cysteine to serine is a relatively conservative mutation and the cysteine is not part of the

dimer interface. This hypothesis could easily be tested by expressing CREB with the mutant zipper domain and assaying for dimerization.

The most interesting of the possible explanations for the failure of the Gag-zipper fusions to dimerize is that the Gag moiety prevents dimerization. This would most likely occur by steric hindrance in which the Gag portion of the protein folds over and occludes the dimerization surface of the zipper. Studies of the HIV-1 Gag protein in solution suggest that Gag domains, rather than assuming an extended “beads on a string” arrangement, may interact with one another in solution, producing just such folded-over conformations (29). While the RSV *in vitro* assembly constructs lack the membrane-binding domain of MA, the C-terminal portion of MA that remains may interact with downstream sequences, or the SP segment may interact with upstream sequences. It might be profitable to repeat the dimerization studies on a CA-SP-LZ fusion in order to examine the role of the upstream MA-p10 sequence. Eliminating the linker sequence or shifting the position of the zipper fusion might also decrease interference, but the effects of doing so are difficult to predict without more information on the conformation of RSV Gag under *in vitro* assembly conditions.

The failure of the FKBP-ligand dimerization system to perform as expected may be due to any of the considerations discussed above or to factors inherent to that system. Only one kit was purchased for these studies, and it is possible that the sample of AP20187 received was defective, or reconstituted improperly, or degraded over the period in which the studies were conducted. Alternatively, the Gag-FKBP fusions may in fact be dimeric without the drug, as suggested by the data presented above on the release of Gag-FKBP from DF1 cells. The FKBP domain appears to be capable of self-

associating to some degree even in the absence of ligand (101), and this ligand-independent dimerization may be potentiated by the presence of upstream Gag sequence. In this scenario, the protein peak on the rate-zonal gradient failed to shift because the ligand-free protein was already dimeric, and the failure to assemble was due to other causes than failure to dimerize. It is also possible that kinetic factors were at fault. The FKBP-ligand dimerization system has been used primarily to induce cell signaling events, which may have very different requirements than does Gag assembly. Kinetics and thermodynamics are decoupled to some extent: if both the association and dissociation rates are high, individual dimers can be very short-lived even among molecules that are dimeric most of the time. It is possible that FKBP-ligand dimerization, while high affinity, does not last long enough to initiate assembly. This hypothesis cannot, however, explain the failure of the protein peak on the rate-zonal gradient to shift in response to the addition of AP20187, as sedimentation should be a purely thermodynamic phenomenon. Overall, the FKBP-ligand system shows less promise for Gag assembly than does leucine zipper-mediated dimerization. Ariad notes that the system has not worked well for the transcriptional regulation, and it may be that Gag assembly is simply another inappropriate application.

Overall, a great deal of work remains to be done to explain the failure to assemble RSV Gag proteins by artificial dimerization. Such experiments are beyond the current scope of my thesis research, but possible future directions will be discussed in Chapter Five.

# CHAPTER FIVE

# PERSPECTIVE

Thus far I have explored several different aspects of the assembly of the Rous Sarcoma Virus Gag protein into virus-like particles. In this chapter I will briefly summarize each of these projects, describe the significance of each to the field of retroviral assembly, and discuss remaining questions and possible directions for future research.

## Chapter 3

- **Summary**: I performed a scanning alanine mutagenesis of the C-terminal 25 amino acids of the p10 domain of Rous Sarcoma Virus Gag and found good correlation between those amino acids required for in vitro assembly and those whose side chains interacted with CA in a published RSV CA extended NTD dimer structure. This mutagenesis allowed me to identify a pair of amino acid residues that when mutated to cysteines would form a disulfide bond across the p10-CA interface. Subsequent cross-linking experiments established that the p10-CA interface depicted in the structure was present in assembled virus-like particles and that the interface formed a hexamer rather than the dimer found in the crystal structure. This information assisted our collaborators in constructing a coarse-grained molecular model of an immature RSV CA NTD hexamer incorporating the p10-CA interface. This model has some striking similarities to electron cryotomography data recently obtained for immature virus-like particles from HIV-1 and RSV.

- **Impact**: This line of research is unique in the retroviral assembly field. The interactions among Gag molecules in the immature particle have been poorly understood relative to those in the mature CA core. This study establishes a mechanism for the previously known importance of the last 25 amino acids of p10 for RSV particle morphology. It confirms the relevance of the published extended NTD structure to retroviral assembly and establishes this structure as the first known example of a protein-protein interaction involving heterologous Gag domains as well as the first and so far only high-resolution structure of a protein-protein interaction unique to the immature VLP. The interaction between p10 and CA is also the best-characterized example of a contribution of a non-CA domain of Gag to protein-protein interactions during assembly (the role of the spacer domain between CA and NC and the significance of the internal “pillars” seen by cryotomography are not yet clear.) Finally, the role of p10 in stabilizing the immature CA NTD hexamer suggests a possible mechanism by which PR cleavage leads to morphological maturation of retroviral particles, the first such mechanism not confined to the β-hairpin formed at the N-terminus of CA upon cleavage. Overall, this work makes a highly significant contribution to our current knowledge of the structure and assembly of immature retroviral particles.

**Chapter 4**

- **Summary**: I attempted to replicate various results in which artificial dimerization of retroviral Gag proteins have been shown to initiate virus-like particle assembly. I employed homodimeric and heterodimeric

leucine zipper fusions, as well as an inducible FKBP dimerization domain, in place of the RSV NC domain in both cell-based and *in vitro* assembly systems. I had only partial success in demonstrating budding of RSV Gag-leucine zipper fusions, previously reported in a baculovirus expression system, from avian cells and was unable to assemble VLPs *in vitro* using any system, although using a highly positively charged zipper in conjunction with nucleic acid produced VLP-like structures with low efficiency. All Gag-leucine zipper fusions tested, as well as the Gag-FKBP fusion, were unable to dimerize *in vitro*.

- **Impact**: This work has not been published and will require a great deal of follow-up if it is to have any impact on the field.

**Future directions**

My study of the role of the p10-CA interaction in RSV Gag assembly is largely complete as it stands; most of the remaining questions pertain to structural questions beyond my current training. For example, it would obviously be desirable to obtain isolated p10-NTD hexamers for high-resolution structural study (full-length Gag is not likely to be amenable to crystallization or to NMR spectroscopy). Two approaches are possible: crystallization of an extended NTD under a variety of conditions in hope of obtaining spontaneous hexamerization, or cleavage of assembled (and possibly cross-linked) particles in hope of having an assembled p10-NTD or p10-CA hexamer left at the end. Neither of these approaches is particularly promising. The first relies on luck, and the second is most likely to be successful if upstream and downstream sequences are removed. This could be accomplished by incorporating the p10-CA cleavage site mutation

described in Chapter 3 and then digesting particles with RSV PR. However, the conditions required for *in vitro* PR digestion may themselves be detrimental to particle structure (see Appendix C). In addition, the CTD should be removed as well to increase the probability of crystallization, and insertion of exogenous retroviral protease sites is problematic as PR recognition sequences are highly variable. These difficulties could be overcome by inserting cleavage sites for a non-retroviral protease, such as TEV protease, into p10 and CA and digesting with that enzyme. Unfortunately, the linker region of CA is exquisitely sensitive to mutation, so insertion of any cleavage site between the NTD and the CTD is not likely to be successful. Finally, it is possible that NTD-CTD interactions or interhexamer NTD-NTD interactions may be required to stabilize the p10-NTD hexamer; interhexamer interactions seen for the helix 4-5 loop of the MLV NTD have been hypothesized to play a role in the immature lattice (87). If so, an isolated, authentic p10-NTD hexamer may be an impossible goal.

In another attempt to obtain detailed structural information on the immature hexamer, I sent samples of several proteins for analysis by two-dimensional crystallography, a technique that has yielded valuable data on the mature HIV-1 lattice structure (44). The structures produced by RSV proteins displayed a persistent curvature, forming either spheres or tubes but no monolayers or crystals under any condition tested. Combined with the "conical" shape of the Gag hexamer (widest at the NTD and narrower beneath, with protein-protein interactions at multiple levels) shown by cryotomography, this suggests the novel possibility that some of the curvature displayed by immature cores is an inherent property of the hexameric lattice rather than a result of pentameric defects. The mathematical difficulties of this idea are

mitigated by the additional finding that the immature lattice is incomplete, with substantial areas of disorganization interspersed with areas of ordered hexamers (130). Unfortunately, for practical purposes, this hypothesis only increases the difficulty of obtaining structural information on the immature hexamer.

A third possibility for structural verification of the p10-CA hexamer model would be to examine the p10-CA interface for evidence of protection for hydrogen-deuterium exchange using mass spectroscopy. This technique has been used successfully to identify a novel NTD-CTD interface in HIV-1 (68) which was later confirmed by 2D crystallography (44) and by assembly interference by a small molecule that binds the CTD side of the interface (Vanda Bartonova, personal communication). Protection of both the p10 and CA sides of the interface in particles relative to unassembled protein would be further evidence of the existence of the interface in particles. It would not directly address whether the interface is truly hexameric, although mass spectroscopy could also be used to verify the size of the disulfide-linked complexes isolated from oxidized CC VLPs. Collaborations on these topics have been discussed and may be forthcoming following the completion of this thesis.

Other remaining questions pertain mostly to peripheral issues. For example, it is not known whether assembly and nuclear export of RSV Gag, which have overlapping amino acid requirements, are actually functionally related. Since the key amino acids of the L-W-V-L NES recognized by the cellular nuclear export factor CRM-1 are also involved in the p10-CA interface, it seems logical that only unassembled Gag can be exported from the nucleus, as assembly would occlude the NES. If this is the case, one might ask whether

the virus or cell has a mechanism (beyond lack of plasma membrane) to prevent nonproductive and potentially toxic nuclear assembly when both Gag and nucleic acid are present in abundance. Alternatively, as nuclear transit of Gag appears to be required for packaging of dimeric RNA, it may be that multiple interacting Gag molecules are required to “collect” the dimeric RNA and export it from the nucleus. To test this hypothesis, one might ask whether an assembly-defective Gag species with an intact NES (such as the p10 T43A mutant) could rescue nuclear export of an NES mutant; unfortunately, since all NES mutants examined to date are also defective for assembly, such an experiment is unlikely to be definitive. The relationship between the assembly and nuclear export functions of the p10 domain of RSV Gag will be difficult to study as the two appear to be inseparable.

Another question is whether cross-linking the p10 domain of Gag to CA would fix the VLP in its immature morphology even after PR cleavage. While conceptually simple, this experiment would be technically difficult for a number of reasons. First, neither the relative timing of oxidation and PR cleavage nor even the timing of either relative to budding is known. The cytoplasm is generally a reducing environment, with disulfide bond formation taking place in the lumen of the endoplasmic reticulum (3). I assume, therefore, that the CC(-11C) Gag particles produced from cells in Chapter 3 became oxidized only after budding and release. PR cleavage appears to take place in the same timeframe. Therefore, simply expressing Gag with both the E51C-T20C double mutation and an active protease and examining the particles produced by thin-section TEM would not be useful, as there would be no way of knowing whether oxidation had taken place prior to PR cleavage. Immature particles could be digested *in trans* by purified PR, as was done in Chapter 3, but

recovering the digested particles for TEM analysis is likely to be challenging, as permeabilizing the membrane sufficiently to allow PR access to the immature Gag core would also allow Gag cleavage products to diffuse away, potentially inhibiting mature core assembly. In addition, digestion *in trans* of sufficient amounts of virus for pelleting and thin-sectioning, as opposed to biochemistry, might be difficult. Overall, while I am confident that cross-linked particles would resist morphological maturation, I fear that the technical challenges to proving so may be insurmountable.

The artificial dimerization project, being less complete, leaves much more scope for future exploration. The most pressing question is why the homodimeric Gag-leucine zipper chimera (∆MBD-bzip) fails to dimerize *in vitro*. While it is possible that the cysteine to serine mutation described in Chapter 4 abrogates dimerization (and this hypothesis would be simple to test), it seems more likely that the Gag moiety of the protein inhibits dimerization, most likely by assuming a conformation that occludes the dimerization surface of the zipper. A possible key to this problem is the observation that the ∆MBD-Kzip mutant protein, which is extremely unlikely to homodimerize due to electrostatic repulsion, produces structures reminiscent of VLPs when mixed with nucleic acid under assembly conditions. This finding hints at an assembly role beyond Gag oligomerization for NC-nucleic acid binding.

Since ∆MBD-Kzip lacks all but the first four amino acids of NC, this role cannot require any nucleic acid-induced conformational change in the NC domain itself but must involve some other part of Gag. The most likely candidate by physical proximity is the upstream CA-SP-NC element (**Figure 4.1**). An interaction between this element and nucleic acid is not

unprecedented; an unpublished NMR structure of the CTD-NC portion of RSV Gag suggests an interaction between SP and the CTD that disappears upon nucleic acid binding (Carol Post, personal communication). The SP domain fits into a hydrophobic groove between helices 1 and 4 of the CTD. The mechanism by which nucleic acid binding displaces SP from the groove is not known.

As the groove itself is conserved among retroviral CA domains (64), the auto-inhibition and release mechanisms may be conserved as well. This begs the question of how artificially dimerized HIV-1 Gag constructs are able to assemble *in vitro* in the absence of nucleic acid. One possibility is that part of the NC domain is required for auto-inhibition. It may be that some NC sequence that is required for auto-inhibition but not for assembly is missing from the artificially dimerized HIV-1 Gag constructs described to date, none of which have included NC sequence. This might also explain why the cell based RSV constructs, which lacked any NC sequence, performed better than the *in vitro* constructs containing the first four amino acids of NC. As discussed in Chapter 4, the presence of the GSGSG linker between the Gag and zipper sequences (**Figure 4.8**) may also have disrupted assembly, as evidenced by the switch from spherical to cylindrical particles in the baculovirus expression system (56) (**Figure 4.2**). However, I consider differences between cellular and *in vitro* assembly systems – such as cellular chaperone activity and the presence of membranes – more likely than the sequence differences to be responsible for the discrepant performances of the leucine zipper fusions in these systems.

While they may not account for the differences between in vitro and cell-based assembly, the potential roles of the CA-SP-NC element and the

GSGSG linker in the failure of the ΔMBD-bzip construct to dimerize *in vitro* require further investigation. I would begin by building a series of constructs in which the mutated zipper domain of ΔMBD-bzip was replaced by a wild-type leucine zipper domain, with and without the GSGSG linker. (Comparison of the CREB and GCN4 zippers might be productive as well; all of the RSV Gag-zipper fusions described in this thesis, as well as those previously published (56), have used the former, while some of the HIV-1 fusions have used the latter (1)). *In vitro* dimerization of a ΔMBD-bzip fusion lacking the cysteine to serine mutation would indicate that the mutation was at fault, while dimerization of a chimera lacking the GSGSG sequence would implicate the linker. If these constructs failed to dimerize, I would truncate the Gag sequence to remove first the remaining N-terminal four amino acids of NC and then the entire SP domain (**Figure 4.1**), again with and without the GSGSG linker. This would indicate whether an auto-inhibitory conformation of the CA-SP-NC element was responsible for the failure to dimerize. As a control, I would also build a full-length ΔMBDΔPR construct with the wild-type zipper domain fused at the end of NC and test for dimerization in the presence and absence of GT8, an eight-base ssDNA oligonucleotide that can bind only one NC domain and does not induce ΔMBDΔPR dimerization (79). Dimerization of this construct only in the presence of GT8 would suggest that the auto-inhibition is relieved by nucleic acid binding as predicted by NMR. These experiments would lay necessary groundwork for understanding artificial Gag dimerization.

The observation that assembly was normal in cells but poor *in vitro* when the NC sequence was interrupted by GSGSG at the NC+4 position suggests that the requirements for the CA-SP-NC element may be more

stringent for *in vitro* assembly. This too requires further investigation, as inclusion of a functional CA-SP-NC element will be required to build a Gag construct that assembles by artificial dimerization. Since the *in vitro* assembly system appears to be more sensitive to disruption of this element than cell-based assembly, I would forego the original strategy of GSGSG insertion and begin with a scanning single or double alanine mutagenesis of the entire CA-SP-NC element (as currently defined in **Figure 4.1**) in the ΔMBDΔPR construct and screen the resulting proteins for assembly *in vitro*. Depending on the results, I would extend this screen further into the NC domain to include all amino acids up to the first zinc finger. This region encompasses 20 amino acids with a net charge of +2 and is not sufficient for *in vitro* assembly without additional basic residues (133), so while the sequence does contain four basic amino acids, I do not anticipate that including any or all of it in a leucine zipper chimera would induce assembly by nucleic acid-mediated oligomerization. The conservative alanine substitution mutagenesis proposed could be followed by insertional mutagenesis as performed in the baculovirus system (60) once the less sensitive sites are identified. These screens should better characterize the CA-SP-NC element for *in vitro* assembly which, combined with additional information on the requirements for Gag dimerization, should suggest a more focused approach to designing an RSV *in vitro* assembly construct driven by artificial dimerization.

Mutagenesis of the CA-SP-NC element in the context of the *in vitro* assembly system would have a second purpose: as a preliminary to further exploration of the relationship of this element to the pillar-like structures observed by cryotomography. Small insertions into less-sensitive parts of the element, or mutation of dispensable amino acids to cysteine followed by

labeling with maleimide-conjugated Undecagold beads (Nanoprobes, Inc.), would be detectable by cryotomography and could demonstrate which parts of the sequence correspond to which parts of the pillars. The collaborator who performed the initial RSV cryotomography study is highly interested in pursuing this line of research if the necessary reagents (purified VLPs assembled from mutant proteins) can be provided. Clarifying the identity and function of these unexpected structures would be a most valuable contribution to the retroviral assembly field.

# APPENDIX A

## Homobifunctional maleimide cross-linking of exogenous cysteine in helix 6 of the CA NTD

Helix 6 of the N-terminal domain of RSV CA forms part of the dimer interface in the extended NTD structure. Unlike the p10-CA portion of the dimer interface, the helix 6 portion of the dimer interface is symmetrical (helix 6 interacts with helix 6). Prior to the alanine mutagenesis of p10, I attempted to demonstrate the existence of the extended NTD interface in virus-like particles by introducing an exogenous cysteine residue into helix 6 and cross-linking using a homobifunctional maleimide reagent, BMH.

This project was undertaken in collaboration with Marc Johnson, then a postdoctoral associate in this laboratory, who selected three candidate amino acids in helix 6 for cysteine mutagenesis and generated the mutations in the context of the pET3xc.p10CANC(-6C) vector. The three amino acids mutated (numbered by position in CA) were D111, G112, and L124.

I expressed and purified these constructs and screened them for assembly *in vitro*. Of the three constructs, only p10-L124C-NC(-6C) was able to assemble (**Figure A.1**). No BMH-mediated cross-linking of the negative control protein, p10CANC(-6C), was observed unless the protein was treated with 6.5 M urea (**Figure A.1**), suggesting that the endogenous cysteine C192, the side chain of which is buried in the RSV CTD structure (64), was not accessible to BMH and would not interfere with the cross-linking assay.

Unfortunately, cross-linking of the p10-L124C-NC(-6C) protein was not specific to VLP assembly. The degree of dimer formation seen was the same whether or not nucleic acid (required for assembly) was added to the

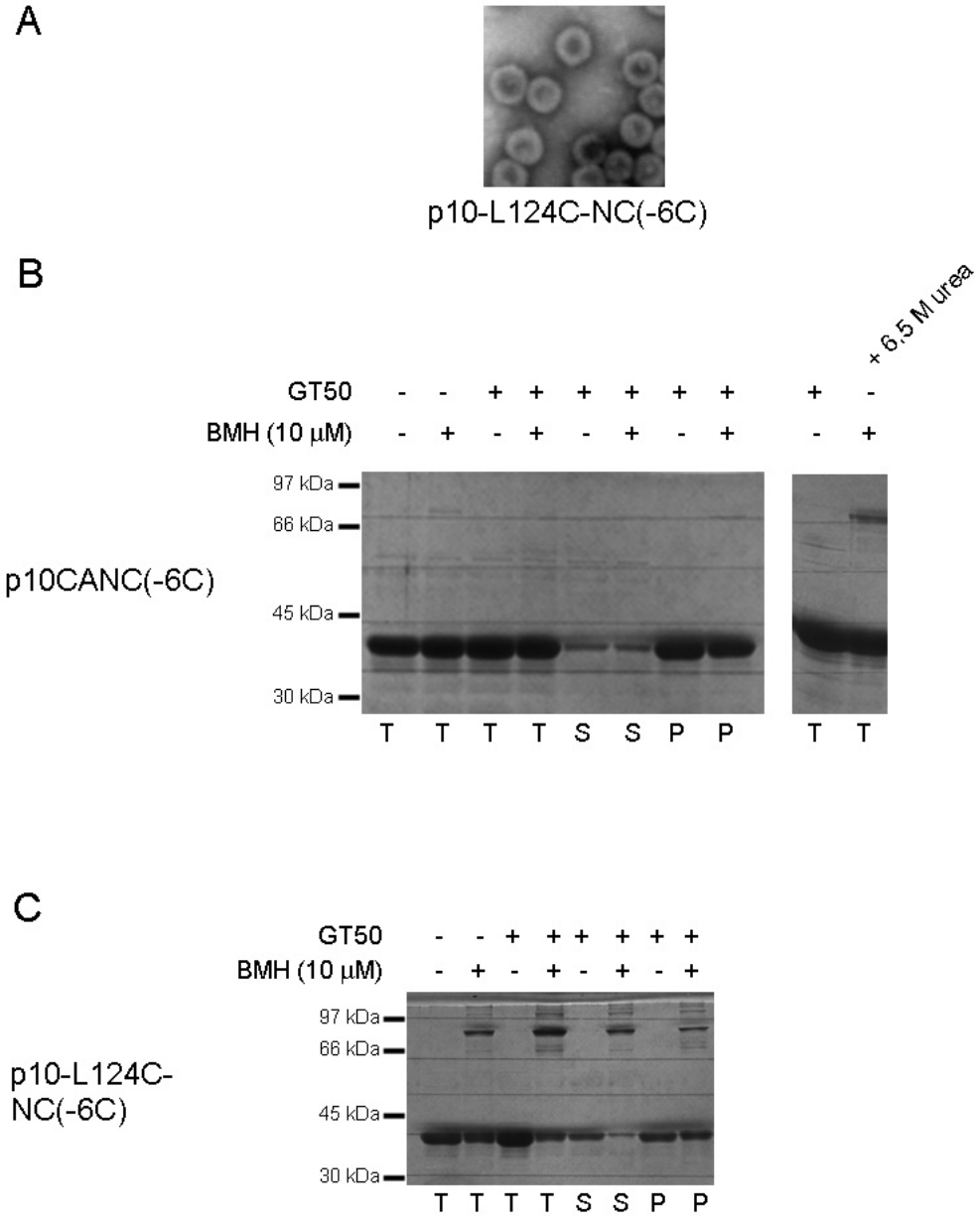



**Figure A.1**: Cross-linking at helix 6 of CA-NTD. **A**. In vitro assembly of p10-L124C-NC(-6C). **B**. Endogenous C192 does not cross-link in the absence of urea. **C**. Cross-linking of exogenous C124 is not specific to assembly. T = total reaction, S = supernatant, P = pellet.

assembly reaction (**Figure A.1**). I also tested for specificity by increasing the concentration of BMH used, on the assumption that kinetics would favor cross-linking of closely apposed cysteines (as in a VLP) even in the presence of excess BMH, whereas non-specific cross-linking of monomeric protein would decrease in the presence of excess BMH. No decrease in dimer formation with increasing BMH was observed for ΔMBDΔPR, suggesting that cross-linking was taking place between the NC or MA-p2 domains within assembled VLPs; however, decreased dimerization with increasing BMH was observed for p10-L124C-NC(-6C) (**Figure A.2**), suggesting that the dimerization observed occurred nonspecifically between molecules randomly colliding in solution. I concluded that the helix 6 interface in the extended NTD structure could not be demonstrated in VLPs by cysteine cross-linking.

The results presented in Chapter Three of this thesis, which were obtained after the experiments described above, suggest several additional considerations. In light of the differences seen in Chapter Three when comparing purified particles (**Figure 3.10**) to total assembly reactions (**Figure 3.6**) and unassembled protein (**Figure 3.9**), it is possible that I would have obtained different results had I purified particles prior to BMH cross-linking. I consider this possibility unlikely, as it appears that the majority of the cross-linking observed was in fact non-specific. Furthermore, the helix 6-helix 6 interaction cannot exist as part of the putative p10-NTD hexamer depicted in **Figure 3.17,** since helix 6 is not part of the NTD-NTD interface. It remains possible that helix 6 of the CA NTD forms inter-hexamer contacts. BMH cross-linking of purified particles might demonstrate just such contacts. However, the presence of inter-hexamer CA-CA contacts would be difficult to prove as p10-CA disulfide cross-linking and CA-CA BMH cross-linking could not be

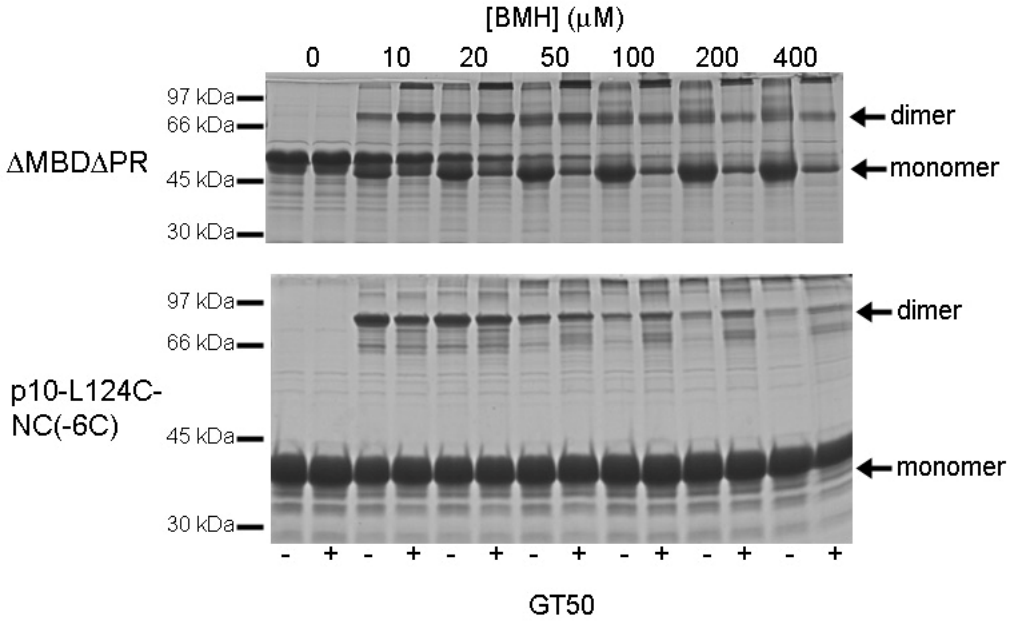



**Figure A.2**: p10-L124C-NC(-6C) dimer crosslinking decreases with increasing concentration of BMH. **Top**: ΔMBDΔPR control. **Bottom**: p10-L124C-NC(-6C). Note decreasing intensity of dimer band from left to right.

performed simultaneously; maleimide-based reagents react only with reduced sulfhydryl groups, and hence the BMH experiments were carried out in the presence of Tris(2-carboxyethyl) phosphine (TCEP), a non-sulfhydryl reducing agent. Overall, I would hypothesize that the helix 6 portion of the extended NTD dimer interface is a crystallization artifact.

# APPENDIX B

## Mutagenesis of solvent-exposed side chains in helix 2 of the CA CTD

May Ma, then a graduate student in this laboratory, had previously hypothesized based on indirect evidence that dimerization of the C-terminal domain of RSV CA is pH-sensitive and occurs much more efficiently at pH 6.5 than at pH 8.0. I hypothesized that this pH-sensitivity might map to the D191 residue of CA, which projects from helix 2 of the CTD analogously to the two amino acids that form the traditional side-by-side HIV CTD dimer interface. The $pK_a$ of the aspartate side chain is ordinarily only 3.9 (25), but side chain $pK_a$ values are dependent on the local peptide environment (i.e., a basic environment will lower the $pK_a$ of an acidic side chain, while a hydrophobic or acidic environment will raise it) so it is possible that the actual $pK_a$ of D191 is much higher. Furthermore, the other side chain that projects from helix 2 is that of R194, suggesting the possibility that these two residues interact across the CTD dimer interface in an electrostatic manner.

In order to test these hypotheses, I made a series of mutations of D191 and R194 and tested the resulting constructs for assembly *in vitro*. Of several single and double mutants tested, only ΔMBD-R194K was able to assemble *in vitro* (**Table B.1**, **Figure B.1**). Subsequent structural studies performed by another group suggest that D191 participates in hydrogen bonding (though not with R194) across a crystallographic CTD dimer interface (Richard Kingston, personal communication), suggesting an alternate hypothesis for the pH-sensitivity of the CTD and the importance of D191 to assembly.

**Table B.1**: CA-CTD helix 2 mutant constructs and *in vitro* assembly results. Amino acid numbering is based on position in CA.

| construct | assembly |
|---|---|
| D191A | - |
| D191E | - |
| D191N | - |
| R194A | - |
| R194K | + |
| D191A-R194K | - |
| D191E-R194K | - |
| D191R-R194D | - |

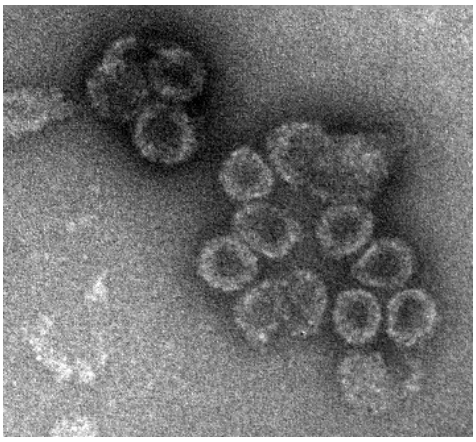


**Figure B.1**: Negative-stained TEM of ΔMBD-R194K particles.

# APPENDIX C

## Disassembly of virus-like particles assembled *in vitro*

Following the p10-CA cross-linking studies, I sought to determine the minimal conditions under which *in vitro*-assembled VLPs would disassemble. Initial results were promising: when 500 mM NaCl or 1 M urea was added to fresh *in vitro* assembly reactions and incubated for one hour, the particles that formed no longer pelleted, suggesting that the particles had disassembled (**Figure C.1**). In addition, the E51C-T20C mutant protein discussed in Chapter 3 (**Figure 3.4**) pelleted in the presence of 500 mM NaCl when oxidized but not when oxidized and then reduced (**Figure C.1**), suggesting that the p10-NTD interface was specifically sensitive to salt.

Unfortunately, these results did not hold up to further study. Purified particles that were pelleted and resuspended in buffer containing 500 mM salt pelleted again during rate-zonal centrifugation, although particles resuspended with 1 M urea did not (**Figure C.2**). The particles resuspended with urea sedimented approximately twice as far as the unassembled protein, consistent with a ΔMBDΔPR hexamer. I interpret this peak as a protein-DNA complex, as the NC-nucleic acid interaction is electrostatic and should not be disrupted by 1 M urea. As the oligonucleotide used, GT50, is expected to bind approximately six molecules of ΔMBDΔPR (the NC binding site size is 8 nt), so it is not possible to say whether any protein-protein interactions contributed to the complex; further experiments with different sized oligonucleotides would be informative.

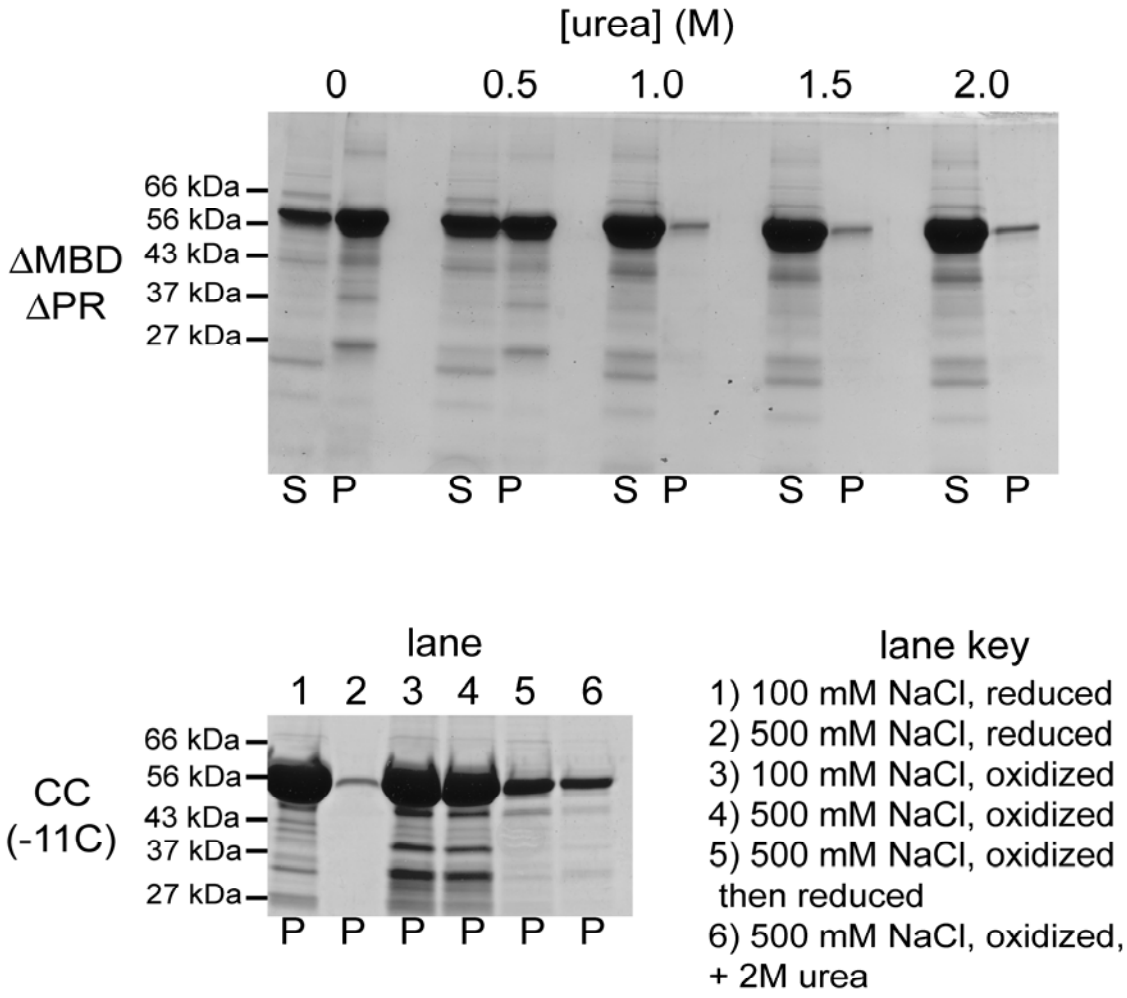



**Figure C.1**: Pelleting of assembly reactions in the presence of urea and salt. **Top**: ΔMBDΔPR assemblies pelleted in the presence of urea. S = supernatant and P = pellet. **Bottom**: CC(-11C) proteins pelleted in the presence of high salt with and without oxidation and added urea. P = pellet; supernatants were not loaded.

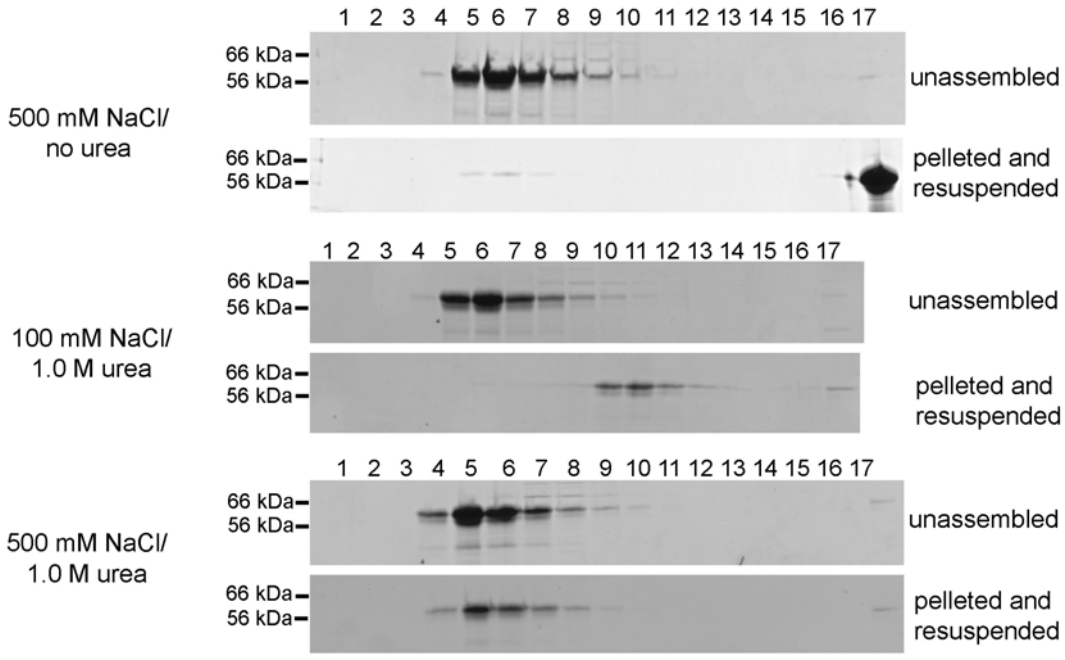



**Figure C.2**: Rate zonal sedimentation analysis of purified ΔMBDΔPR VLPs resuspended with high salt or urea. The centrifugation was performed on 10-30% sucrose gradients made in 50 mM Tris pH 8.0 with salt and urea as noted. Unassembled protein was run under each condtion as a monomeric sedimentation standard. Fraction 17 is the pellet at the bottom of the gradient.

Pelleted particles resuspended with 500 mM NaCl also “reassembled” more efficiently than those resuspended with 1 M urea (**Figure C.3**) upon dialysis back to assembly conditions (50 mM MES pH 6.5, 100 mM NaCl), suggesting that the pelleted particles may not have disassembled after all.

I can think of two hypotheses to explain these discrepancies. First, the initial studies used E51C-T20C(-11C) mutant protein to allow for NTD cross-linking, while the follow-up studies used ΔMBDΔPR. It is possible that there is a difference in the disassembly properties of the two proteins. Second, the purification of the particles in the follow-up studies required an additional four hours for equilibrium sedimentation, and the particles may become resistant to disassembly after this prolonged incubation. Further experiments would be required to test these hypotheses.

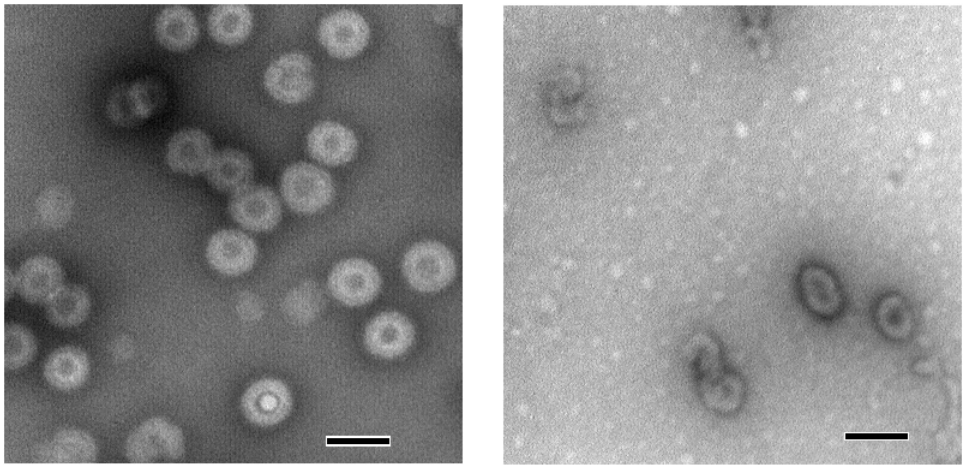



**Figure C.3**: TEM of virus-like particles “reassembled” by dialysis after purification and resuspension under the indicated conditions. Scale bars represent 100 nm.

## REFERENCES


1. **Accola, M. A., B. Strack, and H. G. Gottlinger.** 2000. Efficient particle production by minimal Gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. J Virol **74:**5395-5402.

2. **Ako-Adjei, D., M. C. Johnson, and V. M. Vogt.** 2005. The retroviral capsid domain dictates virion size, morphology, and coassembly of gag into virus-like particles. J Virol **79:**13463-72.

3. **Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter.** 2002. Molecular Biology of the Cell, 4th ed. Garland Science, Taylor & Francis Group, New York.

4. **Alfadhli, A., T. C. Dhenub, A. Still, and E. Barklis.** 2005. Analysis of human immunodeficiency virus type 1 Gag dimerization-induced assembly. J Virol **79:**14498-506.

5. **Alfadhli, A., D. Huseby, E. Kapit, D. Colman, and E. Barklis.** 2007. Human immunodeficiency virus type 1 matrix protein assembles on membranes as a hexamer. J Virol **81:**1472-8.

6. **Amarasinghe, G. K., R. N. De Guzman, R. B. Turner, K. J. Chancellor, Z. R. Wu, and M. F. Summers.** 2000. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. J Mol Biol **301:**491-511.

7. **Birkett, A. J., B. Yelamos, I. Rodriguez-Crespo, F. Gavilanes, and D. L. Peterson.** 1997. Cloning, expression, purification, and characterization of the major core protein (p26) from equine infectious anemia virus. Biochim Biophys Acta **1339:**62-72.

8. **Boeke, J. D., and J. P. Stoye.** 1997. Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements, p. 343-436. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), Retroviruses. Cold Spring Harbor Laboratory Press.

9. **Bouamr, F., C. C. Cornilescu, S. P. Goff, N. Tjandra, and C. A. Carter.** 2005. Structural and dynamics studies of the D54A mutant of human T cell leukemia virus-1 capsid protein. J Biol Chem **280:**6792-801.

10. **Briggs, J. A., M. C. Johnson, M. N. Simon, S. D. Fuller, and V. M. Vogt.** 2006. Cryo-electron microscopy reveals conserved and divergent features of gag packing in immature particles of Rous sarcoma virus and human immunodeficiency virus. J Mol Biol **355:**157-68.

11. **Briggs, J. A., M. N. Simon, I. Gross, H. G. Krausslich, S. D. Fuller, V. M. Vogt, and M. C. Johnson.** 2004. The stoichiometry of Gag protein in HIV-1. Nat Struct Mol Biol **11:**672-5.

12. **Bukrinsky, M. I., S. Haggerty, M. P. Dempsey, N. Sharova, A. Adzhubel, L. Spitz, P. Lewis, D. Goldfarb, M. Emerman, and M. Stevenson.** 1993. A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. Nature **365:**666-9.

13. **Butterfield-Gerson, K. L., L. Z. Scheifele, E. P. Ryan, A. K. Hopper, and L. J. Parent.** 2006. Importin-beta family members mediate alpharetrovirus gag nuclear entry via interactions with matrix and nucleocapsid. J Virol **80:**1798-806.

14. **Callahan, E. M., and J. W. Wills.** 2003. Link between genome packaging and rate of budding for Rous sarcoma virus. J Virol **77:**9388-98.

15. **Campbell, S., R. J. Fisher, E. M. Towler, S. Fox, H. J. Issaq, T. Wolfe, L. R. Phillips, and A. Rein.** 2001. Modulation of HIV-like particle assembly in vitro by inositol phosphates. Proc Natl Acad Sci U S A **98:**10875-10879.

16. **Campbell, S., and A. Rein.** 1999. In vitro assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain. J Virol **73:**2270-2279.

17. **Campbell, S., and V. M. Vogt.** 1997. In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: identification

of the p10 domain as a morphological determinant in the formation of spherical particles. J Virol **71:**4425-4435.

18. **Campbell, S., and V. M. Vogt.** 1995. Self-assembly in vitro of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. J Virol **69:**6487-6497.

19. **Campos-Olivas, R., J. L. Newman, and M. F. Summers.** 2000. Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. J Mol Biol **296:**633-49.

20. **Caspar, D. L., and A. Klug.** 1962. Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol **27:**1-24.

21. **Chen, K., I. Bachtiar, G. Piszczek, F. Bouamr, C. Carter, and N. Tjandra.** 2008. Solution NMR characterizations of oligomerization and dynamics of equine infectious anemia virus matrix protein and its interaction with PIP2. Biochemistry **47:**1928-37.

22. **Choi, G., S. Park, B. Choi, S. Hong, J. Lee, E. Hunter, and S. S. Rhee.** 1999. Identification of a cytoplasmic targeting/retention signal in a retroviral Gag polyprotein. J Virol **73:**5431-7.

23. **Clackson, T., W. Yang, L. W. Rozamus, M. Hatada, J. F. Amara, C. T. Rollins, L. F. Stevenson, S. R. Magari, S. A. Wood, N. L. Courage, X. Lu, F. Cerasoli, Jr., M. Gilman, and D. A. Holt.** 1998. Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. Proc Natl Acad Sci U S A **95:**10437-42.

24. **Copeland, C. S., P. J. Brindley, O. Heyers, S. F. Michael, D. A. Johnston, D. L. Williams, A. C. Ivens, and B. H. Kalinna.** 2003. Boudicca, a retrovirus-like long terminal repeat retrotransposon from the genome of the human blood fluke Schistosoma mansoni. J Virol **77:**6153-66.

25. **Creighton, T. E.** 1993. Proteins: Structures and Molecular Properties, 2nd ed. W.H. Freeman and Company, New York.

26. **Crick, F.** 1970. Central dogma of molecular biology. Nature **227:**561-3.

27. **Crick, F. H., and J. D. Watson.** 1956. Structure of small viruses. Nature **177:**473-5.

28. **Dalton, A. K., D. Ako-Adjei, P. S. Murray, D. Murray, and V. M. Vogt.** 2007. Electrostatic interactions drive membrane association of the human immunodeficiency virus type 1 Gag MA domain. J Virol **81:**6434-45.

29. **Datta, S. A., J. E. Curtis, W. Ratcliff, P. K. Clark, R. M. Crist, J. Lebowitz, S. Krueger, and A. Rein.** 2007. Conformation of the HIV-1 Gag protein in solution. J Mol Biol **365:**812-24.

30. **De Guzman, R. N., Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, and M. F. Summers.** 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. Science **279:**384-8.

31. **Dooher, J. E., and J. R. Lingappa.** 2004. Cell-free systems for capsid assembly of primate lentiviruses from three different lineages. J Med Primatol **33:**272-80.

32. **Dooher, J. E., and J. R. Lingappa.** 2004. Conservation of a stepwise, energy-sensitive pathway involving HP68 for assembly of primate lentivirus capsids in cells. J Virol **78:**1645-56.

33. **Dooher, J. E., B. L. Schneider, J. C. Reed, and J. R. Lingappa.** 2007. Host ABCE1 is at plasma membrane HIV assembly sites and its dissociation from Gag is linked to subsequent events of virus production. Traffic **8:**195-211.

34. **Dryden, K. A., G. Wang, M. Yeager, M. L. Nibert, K. M. Coombs, D. B. Furlong, B. N. Fields, and T. S. Baker.** 1993. Early steps in reovirus infection are associated with dramatic changes in supramolecular structure and protein conformation: analysis of virions and subviral particles by cryoelectron microscopy and image reconstruction. J Cell Biol **122:**1023-41.

35. **Eastman, S. W., and M. L. Linial.** 2001. Identification of a conserved residue of foamy virus Gag required for intracellular capsid assembly. J Virol **75:**6857-64.

36. **Ehrlich, L. S., B. E. Agresta, and C. A. Carter.** 1992. Assembly of recombinant human immunodeficiency virus type 1 capsid protein in vitro. J Virol **66:**4874-83.

37. **Fauquet, C. (ed.).** 2005. Virus taxonomy : classification and nomenclature of viruses ; 8th report of the International Committee on Taxonomy of Viruses, 8th ed. Elsevier/Academic Press, Oxford.

38. **Flint, S. J., Enquist L.W., Krug R.M., Racaniello, V.R., Skalka A.M. (ed.).** 2000. Principles of Virology: Molecular Biology, Pathogenesis, and Control, 1st ed. ASM Press, Washington, D.C.

39. **Fokine, A., P. R. Chipman, P. G. Leiman, V. V. Mesyanzhinov, V. B. Rao, and M. G. Rossmann.** 2004. Molecular architecture of the prolate head of bacteriophage T4. Proc Natl Acad Sci U S A **101:**6003-8.

40. **Forshey, B. M., U. von Schwedler, W. I. Sundquist, and C. Aiken.** 2002. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. J Virol **76:**5667-77.

41. **Fuller, S. D., T. Wilk, B. E. Gowen, H. G. Krausslich, and V. M. Vogt.** 1997. Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. Curr Biol **7:**729-38.

42. **Gamble, T. R., S. Yoo, F. F. Vajdos, U. K. von Schwedler, D. K. Worthylake, H. Wang, J. P. McCutcheon, W. I. Sundquist, and C. P. Hill.** 1997. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. Science **278:**849-53.

43. **Ganser-Pornillos, B. K., A. Cheng, and M. Yeager.** 2007. Structure of Full-length HIV-1 CA: A Model for the Mature Capsid Lattice. Cell **131**.

44. **Ganser-Pornillos, B. K., A. Cheng, and M. Yeager.** 2007. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. Cell **131:**70-9.

45. **Ganser-Pornillos, B. K., U. K. von Schwedler, K. M. Stray, C. Aiken, and W. I. Sundquist.** 2004. Assembly properties of the human immunodeficiency virus type 1 CA protein. J Virol **78:**2545-52.

46. **Ganser, B. K., A. Cheng, W. I. Sundquist, and M. Yeager.** 2003. Three-dimensional structure of the M-MuLV CA protein on a lipid monolayer: a general model for retroviral capsid assembly. Embo J **22:**2886-92.

47. **Ganser, B. K., S. Li, V. Y. Klishko, J. T. Finch, and W. I. Sundquist.** 1999. Assembly and analysis of conical models for the HIV-1 core. Science **283:**80-3.

48. **Grimes, J. M., J. N. Burroughs, P. Gouet, J. M. Diprose, R. Malby, S. Zientara, P. P. Mertens, and D. I. Stuart.** 1998. The atomic structure of the bluetongue virus core. Nature **395:**470-8.

49. **Gross, I., H. Hohenberg, C. Huckhagel, and H. G. Krausslich.** 1998. N-Terminal extension of human immunodeficiency virus capsid protein converts the in vitro assembly phenotype from tubular to spherical particles. J Virol **72:**4798-4810.

50. **Gross, I., H. Hohenberg, T. Wilk, K. Wiegers, M. Grattinger, B. Muller, S. Fuller, and H. G. Krausslich.** 2000. A conformational switch controlling HIV-1 morphogenesis. Embo J **19:**103-113.

51. **Guex, N., and M. C. Peitsch.** 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis **18:**2714-2723.

52. **Huseby, D., R. L. Barklis, A. Alfadhli, and E. Barklis.** 2005. Assembly of human immunodeficiency virus precursor gag proteins. J Biol Chem **280:**17664-70.

53. **Ivanov, D., J. R. Stone, J. L. Maki, T. Collins, and G. Wagner.** 2005. Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain. Mol Cell **17:**137-43.

54. **Ivanov, D., O. V. Tsodikov, J. Kasanov, T. Ellenberger, G. Wagner, and T. Collins.** 2007. Domain-swapped dimerization of the HIV-1 capsid C-terminal domain. Proc Natl Acad Sci U S A **104:**4353-8.

55. **Jasenosky, L. D., G. Neumann, I. Lukashevich, and Y. Kawaoka.** 2001. Ebola virus VP40-induced particle formation and association with the lipid bilayer. J Virol **75:**5205-14.

56. **Johnson, M. C., H. M. Scobie, Y. M. Ma, and V. M. Vogt.** 2002. Nucleic acid-independent retrovirus assembly can be driven by dimerization. J Virol **76:**11177-11185.

57. **Joshi, S. M.** 2000. Analysis of the p10, capsid, and nucleocapsid domains of the Rous sarcoma virus Gag polyprotein during assembly in vitro. PhD. Cornell University, Ithaca.

58. **Joshi, S. M., and V. M. Vogt.** 2000. Role of the Rous sarcoma virus p10 domain in shape determination of Gag virus-like particles assembled in vitro and within Escherichia coli. J Virol **74:**10260-10268.

59. **Keller, P. W.** 2007. Functional analysis of the SP region of Rous sarcoma virus Gag protein in immature assembly. Cornell University, Ithaca.

60. **Keller, P. W., M. C. Johnson, and V. M. Vogt.** 2008. Mutations in SP and adjoining sequences in Rous sarcoma virus Gag lead to tubular budding. J Virol.

61. **Kelley, L. A., R. M. MacCallum, and M. J. Sternberg.** 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol **299:**499-520.

62. **Khorasanizadeh, S., R. Campos-Olivas, and M. F. Summers.** 1999. Solution structure of the capsid protein from the human T-cell leukemia virus type-I. J Mol Biol **291:**491-505.

63. **Kikonyogo, A., F. Bouamr, M. L. Vana, Y. Xiang, A. Aiyar, C. Carter, and J. Leis.** 2001. Proteins related to the Nedd4 family of ubiquitin protein ligases interact with the L domain of Rous sarcoma virus and

are required for gag budding from cells. Proc Natl Acad Sci U S A **98:**11199-204.

64. **Kingston, R. L., T. Fitzon-Ostendorp, E. Z. Eisenmesser, G. W. Schatz, V. M. Vogt, C. B. Post, and M. G. Rossmann.** 2000. Structure and self-association of the Rous sarcoma virus capsid protein. Structure **8:**617-28.

65. **Klikova, M., S. S. Rhee, E. Hunter, and T. Ruml.** 1995. Efficient in vivo and in vitro assembly of retroviral capsids from Gag precursor proteins expressed in bacteria. J Virol **69:**1093-8.

66. **Knejzlik, Z., Z. Smekalova, T. Ruml, and M. Sakalian.** 2007. Multimerization of the p12 domain is necessary for Mason-Pfizer monkey virus Gag assembly in vitro. Virology **365:**260-70.

67. **Kyere, S. K., P. R. Joseph, and M. F. Summers.** 2008. The p12 domain is unstructured in a murine leukemia virus p12-CA(N) Gag construct. PLoS ONE **3:**e1902.

68. **Lanman, J., T. T. Lam, S. Barnes, M. Sakalian, M. R. Emmett, A. G. Marshall, and P. E. Prevelige, Jr.** 2003. Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry. J Mol Biol **325:**759-772.

69. **Larson, D. R., Y. M. Ma, V. M. Vogt, and W. W. Webb.** 2003. Direct measurement of Gag-Gag interaction during retrovirus assembly with FRET and fluorescence correlation spectroscopy. J Cell Biol **162:**1233-44.

70. **Lawton, J. A., M. K. Estes, and B. V. Prasad.** 1997. Three-dimensional visualization of mRNA release from actively transcribing rotavirus particles. Nat Struct Biol **4:**118-21.

71. **Lee, S. K., V. Boyko, and W. S. Hu.** 2007. Capsid is an important determinant for functional complementation of murine leukemia virus and spleen necrosis virus Gag proteins. Virology **360:**388-97.

72. **Li, S., C. P. Hill, W. I. Sundquist, and J. T. Finch.** 2000. Image reconstructions of helical assemblies of the HIV-1 CA protein. Nature **407:**409-13.

73. **Lingappa, J. R., J. E. Dooher, M. A. Newman, P. K. Kiser, and K. C. Klein.** 2006. Basic residues in the nucleocapsid domain of Gag are required for interaction of HIV-1 gag with ABCE1 (HP68), a cellular protein important for HIV-1 capsid assembly. J Biol Chem **281:**3773-84.

74. **Lingappa, J. R., R. L. Hill, M. L. Wong, and R. S. Hegde.** 1997. A multistep, ATP-dependent pathway for assembly of human immunodeficiency virus capsids in a cell-free system. J Cell Biol **136:**567-81.

75. **Lopez-Verges, S., G. Camus, G. Blot, R. Beauvoir, R. Benarous, and C. Berlioz-Torrent.** 2006. Tail-interacting protein TIP47 is a connector between Gag and Env and is required for Env incorporation into HIV-1 virions. Proc Natl Acad Sci U S A **103:**14947-52.

76. **Loriaux, M. M., R. P. Rehfuss, R. G. Brennan, and R. H. Goodman.** 1993. Engineered leucine zippers show that hemiphosphorylated CREB complexes are transcriptionally active. Proc Natl Acad Sci U S A **90:**9046-50.

77. **Ma, Y. M.** 2003. Analysis of Gag-oligonucleotide and Gag-Gag interactions suggests protein dimers as important intermediates of Rous sarcoma virus assembly. PhD. Cornell University, Ithaca.

78. **Ma, Y. M., and V. M. Vogt.** 2004. Nucleic acid binding-induced Gag dimerization in the assembly of Rous sarcoma virus particles in vitro. J Virol **78:**52-60.

79. **Ma, Y. M., and V. M. Vogt.** 2002. Rous sarcoma virus Gag protein-oligonucleotide interaction suggests a critical role for protein dimer formation in assembly. J Virol **76:**5452-5462.

80. **Mayo, K., D. Huseby, J. McDermott, B. Arvidson, L. Finlay, and E. Barklis.** 2003. Retrovirus capsid protein assembly arrangements. J Mol Biol **325:**225-37.

81. **McDermott, J., L. Farrell, R. Ross, and E. Barklis.** 1996. Structural analysis of human immunodeficiency virus type 1 Gag protein interactions, using cysteine-specific reagents. J Virol **70:**5106-14.

82. **McGuffin, L. J., K. Bryson, and D. T. Jones.** 2000. The PSIPRED protein structure prediction server. Bioinformatics **16:**404-5.

83. **Morellet, N., S. Druillennec, C. Lenoir, S. Bouaziz, and B. P. Roques.** 2005. Helical structure determined by NMR of the HIV-1 (345-392)Gag sequence, surrounding p2: implications for particle assembly and RNA packaging. Protein Sci **14:**375-86.

84. **Morikawa, Y., T. Goto, and F. Momose.** 2004. Human immunodeficiency virus type 1 Gag assembly through assembly intermediates. J Biol Chem **279:**31964-72.

85. **Morikawa, Y., T. Goto, and K. Sano.** 1999. In vitro assembly of human immunodeficiency virus type 1 Gag protein. J Biol Chem **274:**27997-8002.

86. **Morita, E., V. Sandrin, H. Y. Chung, S. G. Morham, S. P. Gygi, C. K. Rodesch, and W. I. Sundquist.** 2007. Human ESCRT and ALIX proteins interact with proteins of the midbody and function in cytokinesis. Embo J **26:**4215-27.

87. **Mortuza, G. B., M. P. Dodding, D. C. Goldstone, L. F. Haire, J. P. Stoye, and I. A. Taylor.** 2008. Structure of B-MLV capsid amino-terminal domain reveals key features of viral tropism, gag assembly and core formation. J Mol Biol **376:**1493-508.

88. **Mortuza, G. B., L. F. Haire, A. Stevens, S. J. Smerdon, J. P. Stoye, and I. A. Taylor.** 2004. High-resolution structure of a retroviral capsid hexameric amino-terminal domain. Nature **431:**481-5.

89. **Mothes, W., A. L. Boerger, S. Narayan, J. M. Cunningham, and J. A. Young.** 2000. Retroviral entry mediated by receptor priming and low pH triggering of an envelope glycoprotein. Cell **103:**679-89.

90. **Muller, B., J. Daecke, O. T. Fackler, M. T. Dittmar, H. Zentgraf, and H. G. Krausslich.** 2004. Construction and characterization of a fluorescently labeled infectious human immunodeficiency virus type 1 derivative. J Virol **78:**10803-13.

91. **Nandhagopal, N., A. A. Simpson, M. C. Johnson, A. B. Francisco, G. W. Schatz, M. G. Rossmann, and V. M. Vogt.** 2004. Dimeric Rous sarcoma virus capsid protein structure relevant to immature Gag assembly. J Mol Biol **335:**275-82.

92. **Nelle, T. D., and J. W. Wills.** 1996. A large region within the Rous sarcoma virus matrix protein is dispensable for budding and infectivity. J Virol **70:**2269-76.

93. **Orlinsky, K. J., J. Gu, M. Hoyt, S. Sandmeyer, and T. M. Menees.** 1996. Mutations in the Ty3 major homology region affect multiple steps in Ty3 retrotransposition. J Virol **70:**3440-8.

94. **Ott, D. E., L. V. Coren, E. N. Chertova, T. D. Gagliardi, K. Nagashima, R. C. Sowder, 2nd, D. T. Poon, and R. J. Gorelick.** 2003. Elimination of protease activity restores efficient virion production to a human immunodeficiency virus type 1 nucleocapsid deletion mutant. J Virol **77:**5547-56.

95. **Parent, L. J., R. P. Bennett, R. C. Craven, T. D. Nelle, N. K. Krishna, J. B. Bowzard, C. B. Wilson, B. A. Puffer, R. C. Montelaro, and J. W. Wills.** 1995. Positionally independent and exchangeable late budding functions of the Rous sarcoma virus and human immunodeficiency virus Gag proteins. J Virol **69:**5455-60.

96. **Petrey, D., Z. Xiang, C. L. Tang, L. Xie, M. Gimpelev, T. Mitros, C. S. Soto, S. Goldsmith-Fischman, A. Kernytsky, A. Schlessinger, I. Y. Koh, E. Alexov, and B. Honig.** 2003. Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. Proteins **53 Suppl 6:**430-5.

97. **Pollastri, G., D. Przybylski, B. Rost, and P. Baldi.** 2002. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins **47:**228-35.

98. **Purdy, J. G., J. M. Flanagan, I. J. Ropson, K. E. Rennoll-Bankert, and R. C. Craven.** 2008. Critical role of conserved hydrophobic residues within the major homology region in mature retroviral capsid assembly. J Virol **82:**5951-61.

99. **Raposo, G., M. Moore, D. Innes, R. Leijendekker, A. Leigh-Brown, P. Benaroch, and H. Geuze.** 2002. Human macrophages accumulate HIV-1 particles in MHC II compartments. Traffic **3:**718-29.

100. **Rashkova, S., A. Athanasiadis, and M. L. Pardue.** 2003. Intracellular targeting of Gag proteins of the Drosophila telomeric retrotransposons. J Virol **77:**6376-84.

101. **Rollins, C. T., V. M. Rivera, D. N. Woolfson, T. Keenan, M. Hatada, S. E. Adams, L. J. Andrade, D. Yaeger, M. R. van Schravendijk, D. A. Holt, M. Gilman, and T. Clackson.** 2000. A ligand-reversible dimerization system for controlling protein-protein interactions. Proc Natl Acad Sci U S A **97:**7096-101.

102. **Rumlova-Klikova, M., E. Hunter, M. V. Nermut, I. Pichova, and T. Ruml.** 2000. Analysis of Mason-Pfizer monkey virus Gag domains required for capsid assembly in bacteria: role of the N-terminal proline residue of CA in directing particle shape. J Virol **74:**8452-9.

103. **Saad, J. S., J. Miller, J. Tai, A. Kim, R. H. Ghanam, and M. F. Summers.** 2006. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. Proc Natl Acad Sci U S A **103:**11364-9.

104. **Sakalian, M., S. S. Dittmer, A. D. Gandy, N. D. Rapp, A. Zabransky, and E. Hunter.** 2002. The Mason-Pfizer monkey virus internal scaffold domain enables in vitro assembly of human immunodeficiency virus type 1 Gag. J Virol **76:**10811-20.

105. **Sakalian, M., and E. Hunter.** 1999. Separate assembly and transport domains within the Gag precursor of Mason-Pfizer monkey virus. J Virol **73:**8073-82.

106. **Sakalian, M., S. D. Parker, R. A. Weldon, Jr., and E. Hunter.** 1996. Synthesis and assembly of retrovirus Gag precursors into immature capsids in vitro. J Virol **70:**3706-15.

107. **Sakalian, M., and N. D. Rapp.** 2006. Rescue of internal scaffold-deleted Mason-Pfizer monkey virus particle production by plasma membrane targeting. Virology **345:**317-27.

108. **Sandefur, S., R. M. Smith, V. Varthakavi, and P. Spearman.** 2000. Mapping and characterization of the N-terminal I domain of human immunodeficiency virus type 1 Pr55(Gag). J Virol **74:**7238-49.

109. **Sandefur, S., V. Varthakavi, and P. Spearman.** 1998. The I domain is required for efficient plasma membrane binding of human immunodeficiency virus type 1 Pr55Gag. J Virol **72:**2723-32.

110. **Schatz, G. W., J. Reinking, J. Zippin, L. K. Nicholson, and V. M. Vogt.** 2001. Importance of the N terminus of Rous sarcoma virus protease for structure and enzymatic function. J Virol **75:**4761-70.

111. **Scheifele, L. Z., R. A. Garbitt, J. D. Rhoads, and L. J. Parent.** 2002. Nuclear entry and CRM1-dependent nuclear export of the Rous sarcoma virus Gag polyprotein. Proc Natl Acad Sci U S A **99:**3944-3949.

112. **Scheifele, L. Z., S. P. Kenney, T. M. Cairns, R. C. Craven, and L. J. Parent.** 2007. Overlapping roles of the Rous sarcoma virus Gag p10 domain in nuclear export and virion core morphology. J Virol **81:**10718-28.

113. **Scheifele, L. Z., E. P. Ryan, and L. J. Parent.** 2005. Detailed mapping of the nuclear export signal in the Rous sarcoma virus Gag protein. J Virol **79:**8732-41.

114. **Sfakianos, J. N., R. A. LaCasse, and E. Hunter.** 2003. The M-PMV cytoplasmic targeting-retention signal directs nascent Gag polypeptides to a pericentriolar region of the cell. Traffic **4:**660-70.

115. **Shindyalov, I. N., and P. E. Bourne.** 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng **11:**739-47.

116. **Swanstrom, R., and J. W. Wills.** 1997. Synthesis, Assembly, and Processing of Viral Proteins, p. 263-334. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), Retroviruses. Cold Spring Harbor Laboratory Press.

117. **Tang, C., E. Loeliger, P. Luncsford, I. Kinde, D. Beckett, and M. F. Summers.** 2004. Entropic switch regulates myristate exposure in the HIV-1 matrix protein. Proc Natl Acad Sci U S A **101:**517-22.

118. **Tang, C., Y. Ndassa, and M. F. Summers.** 2002. Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. Nat Struct Biol **9:**537-43.

119. **Telesnitsky, A., and S. P. Goff.** 1997. Reverse Transcriptase and the Generation of Retroviral DNA, p. 121-160. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), Retroviruses. Cold Spring Harbor Laboratory Press.

120. **Ulbrich, P., S. Haubova, M. V. Nermut, E. Hunter, M. Rumlova, and T. Ruml.** 2006. Distinct roles for nucleic acid in in vitro assembly of purified Mason-Pfizer monkey virus CANC proteins. J Virol **80:**7089-99.

121. **Vogt, P. K.** 1997. Historical Introduction to the General Properties of Retroviruses, p. 1-26. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), Retroviruses. Cold Spring Harbor Laboratory Press.

122. **Vogt, V. M.** 1997. Retroviral Virions and Genomes, p. 27-70. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), Retroviruses. Cold Spring Harbor Laboratory Press.

123. **von Schwedler, U. K., T. L. Stemmler, V. Y. Klishko, S. Li, K. H. Albertine, D. R. Davis, and W. I. Sundquist.** 1998. Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. Embo J **17:**1555-1568.

124. **von Schwedler, U. K., K. M. Stray, J. E. Garrus, and W. I. Sundquist.** 2003. Functional surfaces of the human immunodeficiency virus type 1 capsid protein. J Virol **77:**5439-50.

125. **von Schwedler, U. K., M. Stuchell, B. Muller, D. M. Ward, H. Y. Chung, E. Morita, H. E. Wang, T. Davis, G. P. He, D. M. Cimbora, A. Scott, H. G. Krausslich, J. Kaplan, S. G. Morham, and W. I. Sundquist.** 2003. The protein network of HIV budding. Cell **114:**701-13.

126. **Weldon, R. A., Jr., and J. W. Wills.** 1993. Characterization of a small (25-kilodalton) derivative of the Rous sarcoma virus Gag protein competent for particle release. J Virol **67:**5550-61.

127. **Welsch, S., O. T. Keppler, A. Habermann, I. Allespach, J. Krijnse-Locker, and H. G. Krausslich.** 2007. HIV-1 buds predominantly at the plasma membrane of primary human macrophages. PLoS Pathog **3:**e36.

128. **WHO** April 29, 2008 2007, posting date. Global HIV prevalence has levelled off. World Health Organization. [Online.]

129. **Wong, H. C., R. Shin, and N. R. Krishna.** 2008. Solution structure of a double mutant of the carboxy-terminal dimerization domain of the HIV-1 capsid protein. Biochemistry **47:**2289-97.

130. **Wright, E. R., J. B. Schooler, H. J. Ding, C. Kieffer, C. Fillmore, W. I. Sundquist, and G. J. Jensen.** 2007. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. Embo J **26:**2218-26.

131. **Yan, X., P. R. Chipman, T. Castberg, G. Bratbak, and T. S. Baker.** 2005. The marine algal virus PpV01 has an icosahedral capsid with T=219 quasisymmetry. J Virol **79:**9236-43.

132. **Yoder, J. D., T. S. Chen, C. R. Gagnier, S. Vemulapalli, C. S. Maier, and D. E. Hruby.** 2006. Pox proteomics: mass spectrometry analysis and identification of Vaccinia virion proteins. Virol J **3:**10.

133. **Yu, F., S. M. Joshi, Y. M. Ma, R. L. Kingston, M. N. Simon, and V. M. Vogt.** 2001. Characterization of Rous sarcoma virus Gag particles assembled in vitro. J Virol **75:**2753-2764.

134. **Yu, S. F., S. W. Eastman, and M. L. Linial.** 2006. Foamy virus capsid assembly occurs at a pericentriolar region through a cytoplasmic targeting/retention signal in Gag. Traffic **7:**966-77.

135. **Yu, X., Q. C. Yu, T. H. Lee, and M. Essex.** 1992. The C terminus of human immunodeficiency virus type 1 matrix protein is involved in early steps of the virus life cycle. J Virol **66:**5667-70.

136. **Yu, X., X. Yuan, Z. Matsuda, T. H. Lee, and M. Essex.** 1992. The matrix protein of human immunodeficiency virus type 1 is required for incorporation of viral envelope protein into mature virions. J Virol **66:**4966-71.

137. **Zhang, Y., H. Qian, Z. Love, and E. Barklis.** 1998. Analysis of the assembly function of the human immunodeficiency virus type 1 gag protein nucleocapsid domain. J Virol **72:**1782-9.

138. **Zhou, J., R. L. Bean, V. M. Vogt, and M. Summers.** 2007. Solution structure of the Rous sarcoma virus nucleocapsid protein: muPsi RNA packaging signal complex. J Mol Biol **365:**453-67.

139. **Zimmerman, C., K. C. Klein, P. K. Kiser, A. R. Singh, B. L. Firestein, S. C. Riba, and J. R. Lingappa.** 2002. Identification of a host protein essential for assembly of immature HIV-1 capsids. Nature **415:**88-92.

140. **Zlotnick, A.** 2003. Are weak protein-protein interactions the general rule in capsid assembly? Virology **315:**269-74.

141. **Zlotnick, A., S. J. Stahl, P. T. Wingfield, J. F. Conway, N. Cheng, and A. C. Steven.** 1998. Shared motifs of the capsid proteins of hepadnaviruses and retroviruses suggest a common evolutionary origin. FEBS Lett **431:**301-4.